Resolving mammalian relationships with molecules

een wetenschappelijk proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen
volgens besluit van het College van Decanen,
op gezag van de Rector Magnificus,
Prof. Dr. C.M.P.M Blom,
in het openbaar te verdedigen op
vrijdag 10 januari 2003,
des namiddags om 1.30 uur precies

door

Ole Madsen

geboren op 12 september 1965
te Ringkøbing (Denemarken)

Promotor:

Prof. Dr. W.W. de Jong (UvA)

Manuscriptcommissie:

Prof. Dr. J.J. Beintema (RUG)

Dr. J.A.M. Leunissen

ISBN 90-90163665-4

Printed by PrintPartners Ipskamp, Enschede

The research described in this thesis was preformed at the Department of Biochemistry, Faculty of Science, University of Nijmegen, and supported by the TMR program of the European Commission (ERB-FMRX-CT98-0221).

In memories of my parents

Contents

CHAPTER ONE

GENERAL INTRODUCTION

Studying mammals has been of great interest to scientists for many decades. Not because mammals are the most numerous animals; with only about 4138 extant recognized species (table 1; Macdonald 1984) the class of mammals is far less speciose than many other groups of animals, like birds or insects. The reason of this interest rather is that present-day mammals are dominantly present in our environment - having conspicuous appearances and considerable sizes - and are very important from an economical point of view, in husbandry and as pets. Moreover, the interest certainly stems also from the fact that scientists themselves belong to one of the orders of mammals, namely the primates. This thesis describes the gradual progress in resolving the phylogenetic relationships within the most prominent group of mammals, the placental mammals or eutherians, by means of molecular data. It further discusses the molecular evolution of two nuclear-encoded mammalian genes that have been used in our phylogenetic studies: aquaporin 2 and the alpha B2 adrenergic receptor. The following introduction gives some background and explanation concerning phylogenetic terminology and theory, which hopefully makes it easier for the reader with no or little phylogenetic and systematic background to read the subsequent chapters.

What's a mammal?

What's a mammal? This question is not simple to answer in a few words, since mammals display a broad variety of different ecologically adaptive forms (Macdonald 1994). All mammals have three characters not found in any other animals: 1) ***three middle ear bones*** (malleus, incus, and stapes) which are involved in transmitting sound from the outside world to the inner ear; 2) in at least one phase of development they have ***hair*** made up of a protein called keratin; and 3) they have modified sweat glands named ***mammary glands*** in which milk is produced to nurse their newborn youngs. Some other characters are not uniquely found in mammals or not present in all mammals: a lower jaw made up of a single bone (dentary), a heart with four chambers, internal fertilization, and highly differentiated teeth that are replaced just once during an individual's life.

As mentioned, mammals display a great diversity in form and function. The largest mammal (the blue whale) is about 25 meter long and weighs some 120,000 kilogram, the smallest (the Kitti's hog nose bat) weighs 1.5 gram. The largest land mammal (the African elephant) weighs around 6,000 kilogram and is about 7 meter long. Some mammals only feed on one type of diet, such as the giant panda which almost exclusively eats bamboo, whereas others are omnivores (e.g. human, some pig and bear species) eating almost everything that is eatable. Some mammals stay their whole life near

Table 1. A traditional classification of mammals.

Subclass	Infraclass	Order	Some representative species	Families	Species
Protheria					
		Monotremata	Echidnas, Platypus	2	3
Theria					
	Marsupialia				
		Didelphimorpha	Opossums	1	75
		Paucituberculata	Shrew opossums	1	7
		Microbiotheria	El Monito del Montes	1	1
		Dasyuromorphia	Tasmanian devils, Quolls, Dunnarts	3	54
		Peramelemorphia	Bandicoots	2	19
		Notoryctemorphia	Marsupial Moles	1	1
		Diprotodontia	Kangaroos, Wombats, Koalas, Possums	9	109
	Eutheria				
		Artiodactyla	Cows, Pigs, Hippos	10	187
		Carnivora	Cats, Dogs, Bears, Seals	10	264
		Cetacea	Whales, Dolphins	9	76
		Chiroptera	Bats	19	951
		Dermoptera	Flying lemurs	1	2
		Hyracoidea	Hyraxes	1	11
		Insectivora	Moles, Hedgehogs, Tenrecs, Golden moles	6	345
		Lagomorpha	Rabbits, Pikas	2	58
		Macroscelidea	Elephant shrews	1	15
		Perissodactyla	Horses, Tapirs, Rhinos	3	16
		Pholidota	Pangolins	1	7
		Primates	Humans, Apes, Monkeys, Lemurs, Lorises	11	181
		Proboscidea	Elephants	1	2
		Rodentia	Mice, Guinea pigs, Squirrels, Gophers	30	1702
		Scandentia	Tree shrews	1	18
		Sirenia	Sea cows	2	4
		Tubulidentata	Aardvarks	1	1
		Xenarthra	Sloths, Anteaters, Armadillos	4	29

Numbers of families and species from Macdonald (1984).

one location, such as the blind mole rat which never leaves its burrows, whereas a wolf may journey through 1,000 square kilometers during its lifetime. Some mammals live in social groups, whereas others are most of their life solitary. Some mammals have large litter sizes, even up to 32 in the common tenrec, whereas many others only get one young at a time. In some species, e.g. the marsupial mice *Antechinus* and *Phascogale*, males do not even live long enough to witness the birth of their own offspring.

Mammalian systematics

Until recently, the systematics of mammals was exclusively the field of paleontologists and morphologists. The reconstruction of the phylogenetic relationships of mammals by means of morphological characters proved difficult. This is due to their great morphological diversity and complex ecological adaptations, and to the sudden appearance of the ancestors of many of the modern orders in the fossil record some 65 million years ago, indicating a fast and bush-like radiation (see below). Thus, when the first molecular data emerged several mammalian systematists were hopeful, together with their biochemical colleagues, that the unresolved parts of the mammalian tree could be solved by means of molecular data. The enthousiasm of traditional systematists, however, quickly disappeared when the molecular data started to support other relationships than those based on morphological data. The climate between molecular and morphological systematists, not only in mammalian phylogeny, changed to a cold war with accusations and intimidations from both sides (see Felsenstein 2001).

Tabel 2. Eutherian superordinal clades.

Superordinal clade		Orders
Morphological		
	Epitheria	All eutherians except Xenarthra
	Glires	Rodentia, Lagomorpha
	Anagalida	Rodentia, Lagomorpha, Macroscelidea
	Archonta	Scandentia, Primates, Dermoptera, Chiroptera
	Volitantia	Dermoptera, Chiroptera
	Ferungulata	Carnivora, Tubulidentata, Artiodactyla, Cetacea, Perissodactyla, Hyracoidea, Proboscidea, Sirenia
	Ungulata	Tubulidentata, Artiodactyla, Cetacea, Perissodactyla, Hyracoidea, Proboscidea, Sirenia
	Altungulata	Perissodactyla, Hyracoidea, Proboscidea, Sirenia
	Paenungulata	Hyracoidea, Proboscidea, Sirenia
	Tethytheria	Proboscidea, Sirenia
	Ferae	Pholidota, Carnivora
New molecular		
	Afrotheria	Hyracoidea, Proboscidea, Sirenia, Tubulidentata, Macroscelidea, Afrosoricida
	Euarchonta	Scandentia, Primates, Dermoptera,
	Laurasiatheria	Pholidota, Carnivora, Perissodactyla, Cetartiodactyla, Chiroptera, Eulipotyphyla
	Fereuungulata	Pholidota, Carnivora, Perissodactyla, Cetartiodactyla
	Euarchontoglires	Scandentia, Primates, Dermoptera, Rodentia, Lagomorpha

Note that the morphological superordinal clades do not necessarily conform to each other and to figure 1.

Resolving the mammalian tree by means of molecular data turned out not to be an easy task either (de Jong 1998). Therefore, also in the world of molecular systematics it was, and is, not always

peace and love. Molecular data do not always yield congruent results, sometimes caused by systematic errors such as long branches and/or incomplete taxon sampling (see below). And unfortunately the attitude of some molecular systematists was more like "my sequences are longer than your sequences, thus my tree is better than your tree". This led to strong statements about relationships between certain mammals which later turned out to be wrong. Hence, "these are embarrassing times for whom is studying mammalian relationships" (W. de Jong 1998, unpublished).

Below is a brief description of morphologically and molecularly based mammalian trees, with special emphasis on the relationships of the Eutherian orders, as seen by this author in a research field where there are more trees than fit the forest.

Morphological tree

The class of mammals is traditionally classified into some 26 orders (see table 1 for names and representative species). The number of families and species in each order varies tremendously, from some 30 families and 1702 species in Rodentia to one family and one sole species, the aardvark, in Tubulidentata. Of the 26 orders, seven are grouped together in the infraclass Marsupialia, the pouched mammals, and 18 are grouped in the infraclass Eutheria (also called placentals, figure 1). Marsupials and eutherians together form the subclass Theria, to the exclusion of the order Monotremata, the egg-laying mammals, which form the subclass Protheria (table 1). The ordinal relationships amongst eutherians, as presented by Novacek (1992), is displayed in figure 1. Some superordinal groupings, or clades, which have been proposed based on morphology, are shown in this figure and in table 2. The tree is largely based on morphological characters and, although no unanimously supported eutherian tree exists, most morphological systematists will agree with most of the ordinal relationships shown (see e.g. Shoshani and McKenna 1998). This tree, often referred to as the "Novacek tree", has generally been used as a standard reference for molecular systematists, and was in fact one of the primary triggers of this thesis project.

As can be seen, the tree is pretty well resolved. At the base of the tree Xenarthra together with Pholidota are the first to branch off, making them the sister group to all other eutherian orders. Whether Pholidota group with Xenarthra at the base of the tree is morphologically not really clear since others have suggested Pholidota as sister to Carnivora in a clade called Ferae (e.g. McKenna and Bell 1997). Subsequently, a multifurcation leads to several superordinal clades such as Archonta (Primates, Scandentia, Dermoptera, Chiroptera), Anagalida (Rodentia, Lagomorpha, Macroscelidea) and Ungulata (Cetacea, Artiodactyla, Perissodactyla, Hyracoidea, Proboscidea, Sirenia, Tubulidentata). Inside Anagalida the Rodentia and Largomorpha group together to form Glires; in Ungulata the Hyracoidea, Proboscidea and Sirenia form the Paenungulata clade, of which the two latter orders are sisters within Tethytheria.

Molecular tree

In figure 2 one of the best supported current molecular views on mammalian relationships is presented, together with some of the suggested new superordinal names (see also table 2). The relationships of eutherians are derived from a concatenation of 22 mostly nuclear genes (Murphy et al. 2001b), whereas the eutherian outgroup is based on Killian et al. (2001). With two exceptions, namely Artiodactyla and Insectivora, the molecular data support the monophyly (i.e., the sharing of a last common ancestor) of each of the traditionally defined eutherian orders. Strong molecular evidence has

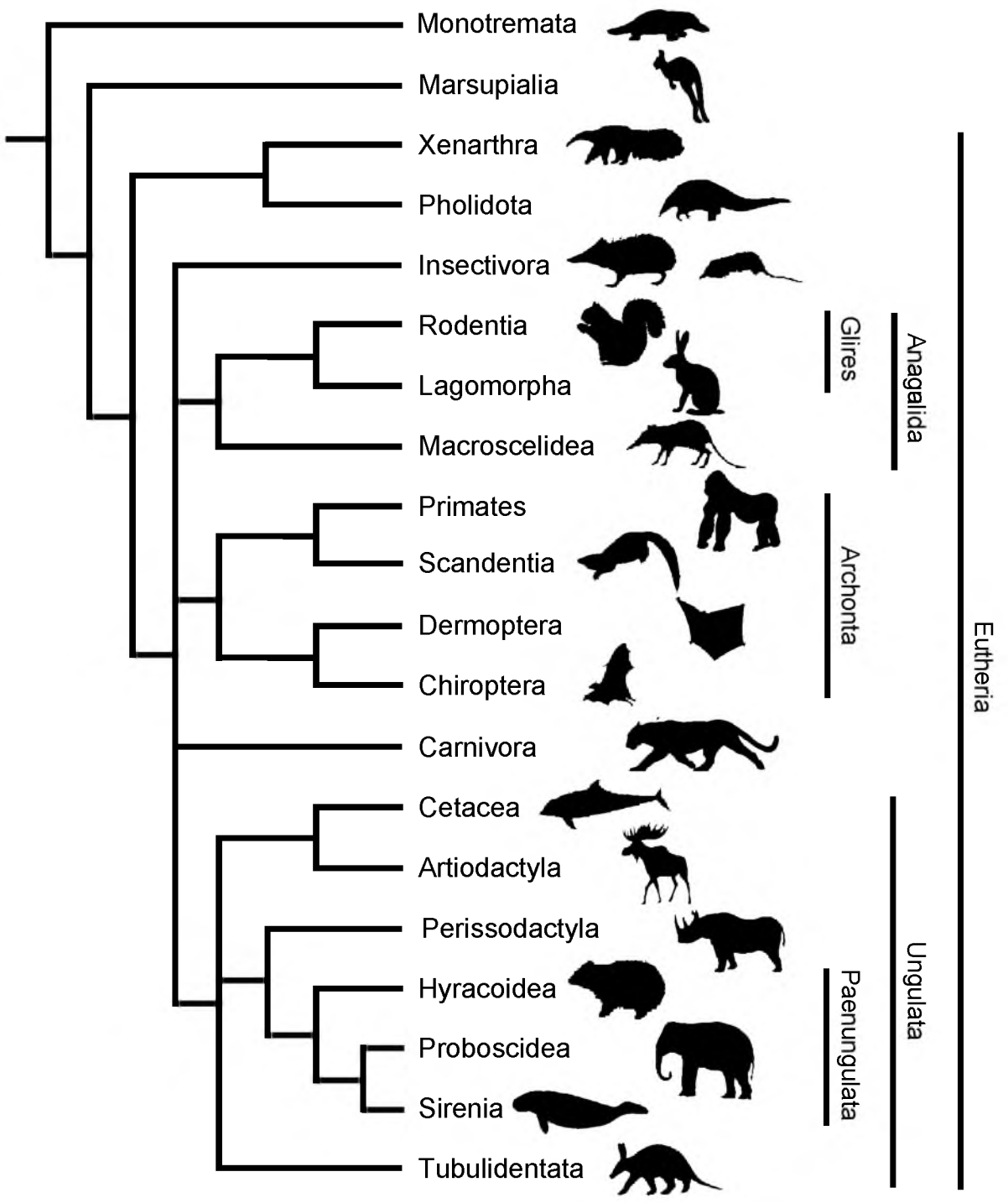


Figure 1. A phylogenetic tree of mammals based on morphological data (Novacek 1992). Different superordinal clades are indicated (see also table 2).

been found for the nesting of Cetacea within the Artiodactyla, as sister group of the hippos (Ursing and Arnason 1998; Gatesy et al. 1999; Kleineidam et al. 1999; Nikaido et al. 1999), joining them into a new molecularly defined order Cetartiodactyla (Montgelard et al. 1997). In contrast, Insectivora has on molecular basis been divided into two separate and unrelated orders: Eulipotyphla (the 'Laurasian' insectivores such as moles, shrews, hedgehogs) and Afrosoricida (the 'African' insectivores such as

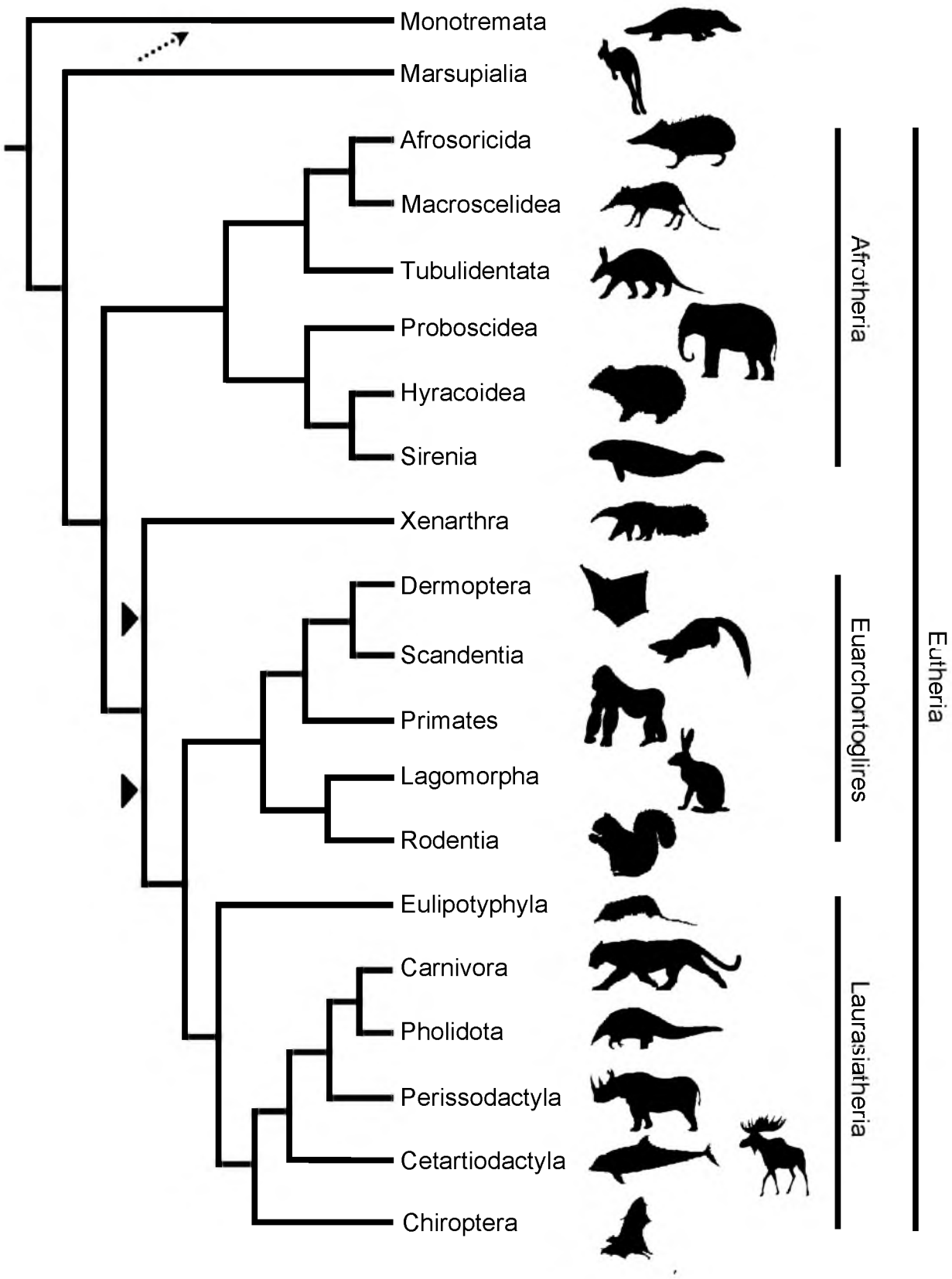


Figure 2. A phylogenetic tree of mammals based on molecular data. The relationships amongst eutherians are based on Murphy et al. (2001b) while those of the outgroups (marsupials and monotremes) are based on e.g. Killian et al. (2001). Arrowheads indicate possible alternative positions of the eutherian root (see text). Dotted arrow indicates the suggested monophyly of monotremes and marsupials (Marsupionta) to the exclusion of eutherians as based on mostly mitochondrial data (Janke et al. 2002). Different superordinal clades are shown (see also table 2).

golden moles and tenrecs) (e.g. Stanhope et al. 1998).

As can be seen from figure 2, molecules have thoroughly remodeled the mammalian tree. Recent analyses on large concatenated mostly nuclear data sets strongly support the division of Eutheria into four groups: Afrotheria (Sirenia, Hyracoidea, Proboscidea, Tubulidentata, Macroscelidea, Afrosoricida); Xenarthra; Euarchontoglires (Dermoptera, Scandentia, Primates, Lagomorpha, Rodentia); and Laurasiatheria (Eulipotyphla, Carnivora, Pholidota, Perissodactyla, Cetartiodactyla, Chiroptera) (Eizirik et al. 2001; Madsen et al. 2001; Murphy et al. 2001a,b). Of the four major clades Afrotheria is most probably basal, followed by Xenarthra, and with Euarchontoglires plus Laurasiatheria as sister clades forming Boreoeutheria (Springer and de Jong 2001). These molecular data strongly support the traditionally recognized superordinal groups Paenungulata and Glires. Another large dataset extensively used in eutherian phylogeny is that of 12 protein-coding genes of the mitochondrial genome (e.g. Pumo et al. 1998; Reyes et al. 2000; Schmitz et al. 2000; Nikaido et al. 2001; Lin and Penny 2001; Arnason et al. 2002). The mitochondrially based mammalian trees conform more or less with those based on nuclear data, but there are important differences. Most importantly, at the very base of the mammalian tree, mitochondrial data have suggested that monotremes and marsupials together form a monophyletic group (Marsupionta) to the exclusion of Eutheria (Janke et al. 2002). Since mitochondrial sequences may not be as robust as nuclear genes in resolving deep mammalian relationships (Springer et al. 2001, see below) this finding should be taken with great caution. Indeed, most nuclear-based trees dealing with this question support the traditional view with eutherians and marsupials together forming Theria (e.g. Killian et al. 2001).

Mitochondrial data have further suggested that rodents are not monophyletic, making some rodents like Guinea pig and squirrel more closely related to other eutherians than to mouse and rat (D'Erchia et al. 1996; Reyes et al. 2001). But as the number of sequenced rodent mitochondrial genomes has increased, and phylogenetic analyses improved, the proposed diphyly of rodents has been weakened and monophyly eventually restored (Cao et al. 2000). Another discrepancy between mitochondrial and nuclear trees is the division of the insectivores. Like the nuclear genes, mitochondrial data also suggest the splitting of the traditional order Insectivora. However, where nuclear data indicate two separate insectivore groups (see above), mitochondrial data divide insectivores into at least three unrelated groups: 1) moles and shrews, placed in Laurasiatheria; 2) golden moles and tenrecs, placed in Afrotheria, and 3) hedgehogs, placed at the very base of the eutherian tree (Krettek et al. 1995; Mouchaty et al. 2000). The distinct placement of hedgehog has been suggested to be due to the nucleotide composition of the hedgehog mitochondrial genome, which deviates significantly from that of other mammals, and thus is likely to influence its position in the tree.

The proposed basal position of the hedgehog immediately illustrates another disagreement between nuclear and mitochondrial eutherian phylogenies, namely the rooting. As mentioned, nuclear data indicate Afrotheria with high confidence as basal (Murphy et al. 2001b), but the root might also be located on the branch to Xenarthra or between Afrotheria/Xenarthra and Laurasiatheria/ Euarchontoglires (Eizirik et al. 2001; Madsen et al. 2001; Murphy et al. 2001a) (arrowheads in fig. 2). However, placing the root within any of the four major clades is statistically rejected in all nuclear trees. In contrast, when hedgehog is left out of the mitochondrial tree, because of its deviating base composition, the eutherian root is placed inside Euarchontoglires, most frequently with rodents at the base (e.g. Reyes et al. 2000)

It thus appears that mammalian phylogeny is a fascinating research field with a lot of answers and even more questions.

Eutherian divergence times

To understand the mode of evolution of eutherians with respect to their biogeographical distribution and early dispersal, it is necessary to know when the different orders of eutherians diverged. In the past years there has been quite some debate between paleontologists and molecular systematists concerning the dating of these divergences. Before the first molecular estimates of eutherian divergence times saw the daylight, a common view was - and probably still is - that the eutherian orders diverged in a rapid and bushlike way around the Cretaceous-Tertiary (K/T) boundary, some 65 million years (Myr) ago. At that time a remarkable mass extinction of dinosaurs and other animals happened, caused by a fast global climate change from a mild, warm and stable to a cold and varied environment. The trigger of this global climate change was probably an asteroid colliding with our Earth, by which a massive amount of dust was thrown into the atmosphere and caused the 'observed' climatic changes (Alvarez et al. 1980). Once dinosaurs went extinct, the already existing small eutherians could take over the ecological niches left open by the dinosaurs and rapidly diversify into a variety of different ecologically adapted forms.

This diversification scenario implies that only a few different eutherian lineages were present before the K/T boundary, and that the diverge times of most of the extant orders should be around that same age (Easteal 1999). However, recent molecular studies have suggested much earlier divergences of the eutherians, beginning as far back as ~100 to ~130 Myr ago, and with all superordinal clades and most eutherian orders already having diverged before the K/T boundary (e.g. Kumar and Hedges, 1998; Bromham et al. 1999; Penny et al. 1999; Eizirik et al. 2001; Madsen et al. 2001; Murphy et al. 2001b). The split between eutherians and marsupials has been found to be around 170 Myr ago (Kumar and Hedges 1998; Penny et al. 1999), which is in good agreement with paleontological estimates. Thus, molecular datings imply that the ancestors of most extant eutherian orders were already present at the time when the dinosaurs became extinct, and that a gap of at least 17 Myr exists in the fossil record (Hedges and Kumar 1999). It has been suggested that such a gap could be explained by the relatively poor fossil record on the southern hemisphere continents (the "Garden of eden" hypothesis) (Foote et al. 1999), where eutherians likely have their origin (Eizirik et al. 2001; Madsen et al. 2001; Murphy et al. 2001a,b). Alternatively, early primitive eutherians from the Cretaceous may actually represent the ancestors of extant lineages, but have not been recognized as such since "they don't look like their modern descendants because they had not yet experienced an adaptive radiation" (Easteal 1999).

Geological time scale and plate tectonic movements

In phylogenetic studies, geological periods are often used to indicate the age of a fossil or the time of divergence. Thus, instead of using exact time points, expressions like "early Cretaceous" or "late Eocene" are sometimes preferred. In figure 3A the geological time scale is shown with some commonly used eras and periods indicated.

With reliable eutherian divergence times and phylogenies in hand, these can be related to the

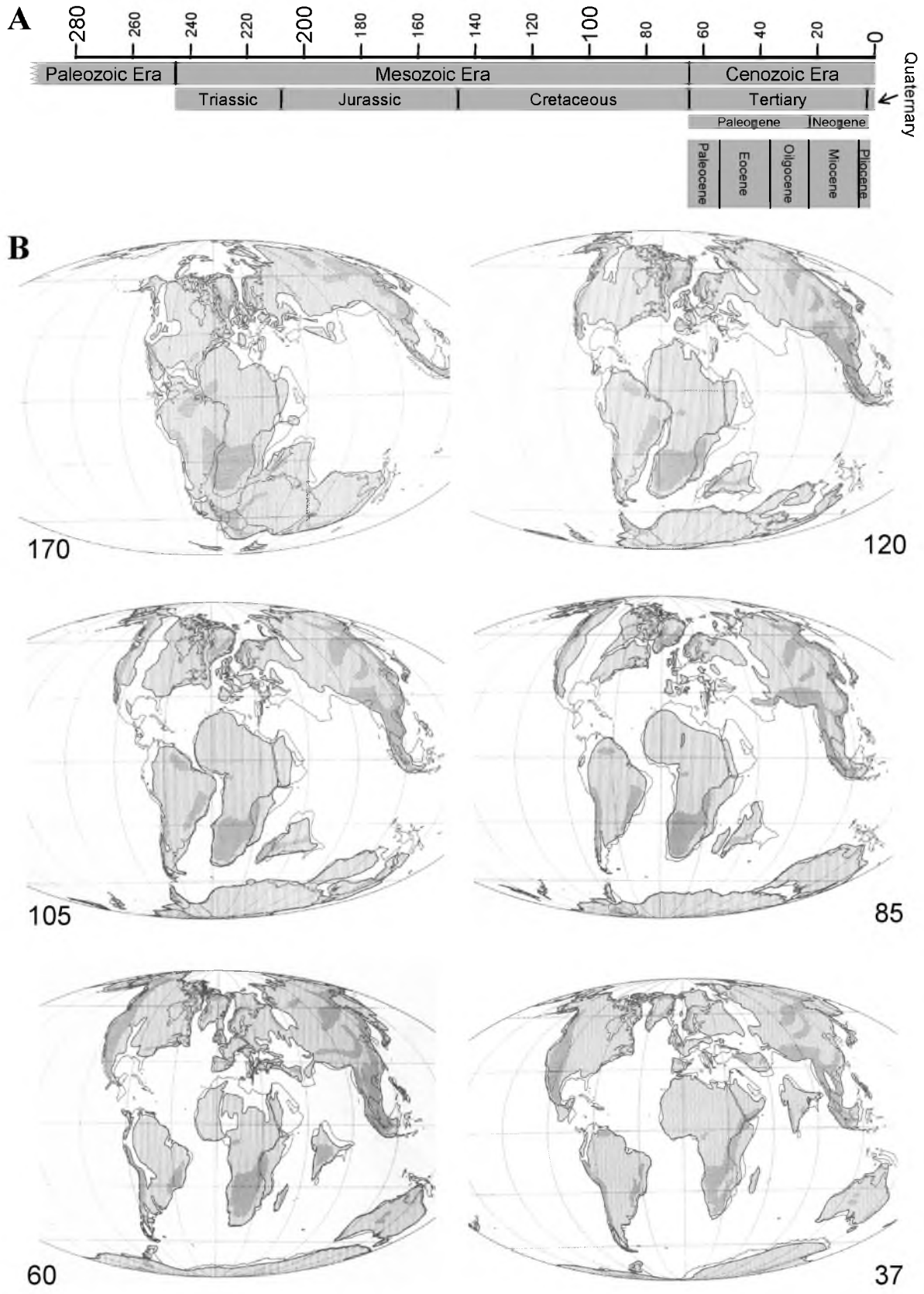


Figure 3. A) Geological timescale (in million years). B) Paleocostline maps at the indicated time periods (in million years ago). Gray areas depict the paleo-continents, whereas present-day coastlines are indicated with lines. Maps taken from Smith et al. (1984).

geographical distribution of fossil finds, and to the geological evidence of plate tectonic. In recent years it has become increasingly evident that continental drift and subsequent isolation may have played an important role in the distribution and initial radiation of early eutherian mammals (e.g. Hedges et al. 1996; Springer et al. 1997; Madsen et al. 2001; Murphy et al. 2001a,b). Although the idea of continental drift is quite old - the Dutch map maker Abraham Ortelius was the first to recognize the fit of South American and African coastal lines in 1596 - the scientific knowledge about plate tectonics is quite new, leaving space for different theories and interpretations of the exact timing of continental fissions and fusions.

An interpretation, largely based on www.scotese.com/earth.htm and Smith et al. (1994), is represented in figure 3B. At the time when marsupials and eutherians diverged, some 170 Myr ago, all continents were still connected in the supercontinent Pangea. Pangea had been formed during the mid and late Paleozoic, and started to break up in the early Jurassic (about 225 to 200 Myr ago). When marsupials and eutherians diverged it was probably still possible for land animals to migrate from pole to pole. In the middle Jurassic, Pangea broke into two supercontinents: the southern hemisphere Gondwana (South America, Africa, India, Australia, Madagascar and Antarctica) and the Northern hemisphere Laurasia (North America, Europe, Asia, Greenland and Iceland). Simultaneously, Gondwana started to drift apart, but is was not until mid Cretaceous before Africa drifted away from South America (~105 Myr ago). Africa became isolated until the early Cenozoic when is got connected with Europe. Also in mid Cretaceous India with Madagascar separated from Antarctica. In the late Cretaceous, India and Madagascar split, and Madagascar drifted northwards parallel with Africa. India moved towards Asia and started to collide in the early Cenozoic, some 50 Myr ago, which resulted in the Tibetan plateau and Himalayas. The final split of the Gondwana continent occurred in the early Cenozoic, when Australia rifted away from Antarctica and started to move rapidly to the north. Laurasia stayed as one landmass during most of the Mesozoic period. In the Cretaceous, North America with Greenland started to rift apart from Europe, opening up the North Atlantic, and in the early to mid Cenozoic North America and Greenland had completely split away from Europe. Thus, until al least 60 to 50 Myr ago animals could spread time and again over the laurasian continent.

Choice of data types for phylogenetic analyses

Before starting a phylogenetic study of any given group of organisms a choice must be made of the type of data one would like to use. One should ask oneself the question: which type of data is potentially best suited to give me an answer to my problem, and is it technically possible to obtain these data? What kind of data can be used for phylogenetic problems? The most commonly used data are morphological and molecular characters, but in fact any character which is inherited from one generation to another can be used in systematic studies provided that it has a reasonable rate of change over time.

For molecular data one requirement must absolutely be fulfilled to be useful in phylogenetic studies of species: the sequences to be compared must not only be homologous, but also orthologous. That is, the sequences must be descended from a common ancestor, and have diverged as a result of speciation events. As a consequence, orthologous genes and proteins usually perform exactly the same function in each taxon. In contrast, genes that are homologous as a result of gene duplication, called

paralogous, may frustrate phylogenetic analyses if the duplication has occurred before the species to be investigated have diverged. But any molecular data that meet this prerequisite of orthology can be used for phylogenetic analyses.

However, other aspects are also important to keep in mind. Firstly, it should of course be technically feasible to obtain the orthologous sequences from different species. In our case this means that it must be possible to determine them after amplification from genomic DNA by the Polymerase Chain Reaction (PCR). Since many of the taxa used in our studies are rare and/or endangered, it is not possible to get fresh material for isolation of mRNA and sequencing the cDNA. In such case, genomic DNA is the only available source of genetic material. In this thesis the coding parts of nuclear genes are the preferred sequences. This implies that the genes of interest should be intronless in their coding region or have 'large' exons. Another aspect is that the content of GC bases should not be too high since this will make PCR difficult or impossible. In addition, a GC skew could also cause errors in the phylogenetic inferences, since not all methods for phylogenetic tree constructions can deal with such GC bias (see below). Secondly, the genes chosen for phylogenetic studies in this thesis should also be of interest from a functional and evolutionary point of view. These aspects in themselves are not important for resolving phylogenetic problems, but by selecting genes coding for e.g. transcription factors, growth factors and receptors, one can kill two birds with one stone: gain novel information about the phylogenetic relations of species and at the same time about e.g. important regions in the proteins encoded by the genes that are compared.

Gene sequence data can be of mitochondrial or nuclear origin. Both types of data have their advantages and disadvantages. Some advantages for choosing mitochondrial genes are that they are without any doubt orthologous and that recombination does not take place, or only extremely rarely (see Ladoukakis and Zouros 2001). Thus the gene genealogy is likely to follow the species phylogeny. Furthermore, it is technically relatively easy to determine mitochondrial gene sequences since all genes are in close proximity in a small genome, which makes direct cloning or PCR amplification more practicable. A disadvantage of using protein-coding mitochondrial genes is that they can be considered as 'a single cistron' with all genes taking part in the same cellular process and effectively forming one transcription unit. Thus, the different genes do not evolve independently of each other, and consequently mutations in one gene may trigger correlated parallel mutations in other genes which eventually may result in a wrong tree. Another disadvantage is that the rate of DNA substitution in the mitochondria may not be suitable for the phylogenetic question to be answered. It has been shown that most nuclear genes used in mammalian phylogeny perform much better per nucleotide than mitochondrial genes in resolving deep mammalian relationships (Springer et al. 2001).

A drawback of employing nuclear genes is the potential risk of using sequences of paralogous genes or pseudogenes as if they were orthologous. Such mistake obviously can lead to wrong inferences about phylogenetic relationships. However, in most cases the error will be detected since the placement of the paralogous gene in the tree will be 'too odd' to be true. Another problem with the many large gene families present in the nuclear genome is that evolutionary processes such as recombination, unequal crossing-over and gene conversion may blur the homologous or orthologous origin of a gene. The greatest advantage of using nuclear genes is that there are so many of them. Hence, it will always be possible to find genes which have the appropriate rates of substitution for the phylogenetic question in mind.

Methods for phylogenetic tree construction

After collecting the sequence data they must be analyzed in order to determine the relationships of the species in question. The first step before performing any phylogenetic analyses is to create a multiple alignment of the sequences. In a multiple sequence alignment a matrix of the data is made, where the individual columns represent the character states at every homologous position in the sequences. The objective of any method for phylogenetic tree construction is to find the tree that fits the data best under a given optimality criterion. Three types of optimality criteria are most often used to calculate phylogenetic relationships: maximum parsimony, minimum distance - also called minimum evolution - and maximum likelihood. Maximum parsimony and maximum likelihood are so-called character-based methods, which means that each site in an alignment is considered as an independent character. Minimum distance is, as the name indicates, a distance-based method relying on the genetic distance between pairs of sequences.

Maximum parsimony

Under the maximum parsimony criterion the best tree is the one that minimizes the number of evolutionary changes given the data (Fitch 1971). The theorem of parsimony is that simpler solutions are being preferred over more complex ones, thus parsimony favors hypotheses that maximize congruence and minimize homoplasy (e.g., parallel, convergent, or superimposed changes). Maximum parsimony methods rely on the detection of phylogenetically informative sites, which are sites where one tree is preferred over others.

Minimum evolution

In minimum distance methods the pairwise evolutionary distances are estimated, as based on a pre-selected model of sequence evolution, and the tree where the sum of all branches is the shortest is constructed. For DNA sequences several models of sequence evolution have been developed (table 3). These models vary from the simple Jukes and Cantor (JC) model, which assumes that nucleotide frequencies are equal and that all possible substitutions equally likely (Jukes and Cantor 1969), to the complex General Time Reversal (GTR) model, which assumes that base frequencies are unequal and that each type of nucleotide substitution has its own rate (six rate parameters) (Lanave et al. 1984; Rodriquez et al. 1990; Yang 1994).

The first nine DNA models in table 3 (JC to GTR) are all restricted to follow a stationary homogeneous Markov process. That means that 1) the rates of changes are constant throughout the tree, and 2) base frequencies remain constant over time. In particular the assumption that base frequencies are constant between taxa may easily be violated, since lineage-specific base compositional bias (G-C bias) can occur as a consequence of e.g. different codon usage between species. To deal with such cases, where non-stationarity is observed, the LogDet/paralinear model of sequence evolution has been developed (Lake 1994; Lockhart et al. 1994; Steel 1994).

Exchangeability matrices for amino acid replacements have also been developed. In these matrices the probability of a change from one amino acid to an other has empirically been determined by simply counting the number of amino acid replacements in closely related proteins. Some of the first amino acid replacement matrices (the PAM matrices) have been developed by Dayhoff et al. (1978), which later were updated by Jones et al. (1992) in the so-called JTT matrix.

Table 3. Different models of nucleotide sequence evolution and model parameters.

Model of sequence evolution	
JC, (Jukes and Cantor 1969)	equal base frequencies, one substitution type
F81, (Felsenstein 1981)	unequal base frequencies, one substitution type
TAJNEI, (Tajima and Nei 1984)	unequal base frequencies, one substitution type
K2P, (Kimura 1980)	equal base frequencies, unequal transition/transversion (Ti:Ts)
F84, (Felsenstein 1984)	unequal base frequencies, unequal Ti:Tv
HKY85, (Hasegawa Kishino and Yano 1985a)	unequal base frequencies, unequal Ti:Tv
K3P, (Kimura1981)	equal base frequencies, 3 substitution types
TAMNEI, (Tamura and Nei 1993)	unequal base frequencies, 3 substitution types
GTR, (Lanave et al. 1984; Rodriquez et al. 1990; Yang 1994)	unequal base frequencies, 6 substitution types
LOGDET, (Steel 1994; Lake 1994; Lockhart et al. 1994)	unequal base frequencies between sequence pair
Model parameters	
Gamma rate (Γ)	rate heterogeneity among sites
p-invar (I)	fraction of invariable sites

General assumptions used by all models of sequence evolution are that all sites can evolve and that every site evolves with the same rate. For protein-coding DNA this is certainly not the case, since third codon positions evolve faster than first and second codon positions - meaning rate heterogeneity among sites - and protein structural and functional constraints imply that some sites are not allowed to change. Two parameters are usually incorporated into the models of sequence evolution to deal with these assumptions: a discrete gamma (Γ) distribution with four to eight rate categories is often used to describe rate heterogeneity among sites (Yang 1993), whereas an invariable-sites model (I) is used to describe sites incapable of accepting changes (e.g., Hasegawa et al 1985b). Incorporation of these model parameters (Γ and/or I) in the different models of sequence evolution gives a fairly well description of most molecular data. It should be noted that the Γ parameter can not be combined with the LogDet model of sequence evolution.

Maximum likelihood

A maximum likelihood (ML) method of phylogenetic inference evaluates the probability that a proposed model of sequence evolution gives rise to the data which is observed given the evolutionary history, i.e. the tree. Thus, given a model of sequence evolution and the data (e.g., the DNA alignment), the tree with the highest likelihood (the ML tree) can be estimated. As mentioned, the ML method is a character based method; the likelihood of a tree, as based on the data, is therefore the sum of the likelihoods of each site. This implies that each site is considered independently with respect to the tree. The models of sequence evolution and the model parameters that can be used in ML tree inference are the same as for minimum distance methods, except for the LogDet model (Table 3). The use of the ML method to estimate phylogeny from molecular data was first proposed by Cavalli-Sforza and Edwards (1967), for gene frequency data, but at that time the method was computationally too difficult to implement in a computer program. Later, algorithms were developed to construct ML phylogenetic trees from DNA sequence data (Felsenstein 1981; Adachi and Hasegawa 1995; Yang 1997; Swofford 1998) and protein sequence data (Kishino et al. 1990; Adachi and Hasegawa 1995; Yang 1997; Felsenstein 2001).

Which model of sequence evolution to use?

The 64000 $ question is of course: which model of sequence evolution should I use for my data? Or put in an other way, which model of sequence evolution is fitting my data the best? A powerful way to compare two models of DNA sequence evolution is by using a likelihood ratio test (LRT) (Goldman 1993; Yang et al. 1994; Huelsenbeck and Rannala 1997). The LRT is in fact one of the cornerstones in ML statistics, and is a solid statistical test by which competing hypotheses (e.g., two different models) are compared by employing a statistic which is based on the ratio of the likelihoods under the two competing hypothesis. If the two hypotheses are nested - which is the case for the different models in table 3 because all models are special cases of the GTR model - the LTR statistic will approximate a Chi-Square distribution. Thus, in these cases the Chi-Square distribution can be used as the critical value to judge if there is significant difference between the two models given a chosen significance value, e.g. 95%.

Which method to use?

Each of the methods for phylogeny reconstruction has its own advantages and disadvantages. Morphologists generally advocate maximum parsimony, whereas in recent years maximum likelihood methods have received ever more support from theoretical and molecular systematists. Unfortunately, it is often necessary to apply all commonly used methods in order to satisfy any potential reviewers who might prefer an other method than oneself. A thorough discussion of the pro's and con's of the different methods is beyond the scope of this introduction, but a few considerations are worth mentioning.

The parsimony criterion, selecting the tree(s) that require the fewest evolutionary changes, is often not fitting for molecular data. Only in cases where the number of changes per site is small will parsimony be as robust as ML in finding the correct tree (Goldman 1990; Steel and Penny 2000). When sequences become more diverse, the level of homoplasy (e.g., parallel, convergent or superimposed changes) increases, and the parsimony criterion that fewest evolutionary changes can describe the true tree becomes increasingly unrealistic. Moreover, in cases where the true tree has long terminal branches connected by short internal branches (the so-called Felsenstein zone), parsimony especially tends to group the long branches together, a phenomenon called long branch attraction, and thus resulting in a wrong tree (Felsenstein 1978; Huelsenbeck 1997).

The ML method is much less sensitive to long branch attraction than parsimony, since models used by the ML method can take superimposed changes into account. In fact, ML estimates of phylogeny will always converge to the true topology, as long as sufficient amounts of data are used and the model of sequence evolution fits the data (Chang 1996; Swofford et al. 1996; Rogers 1997). Another attractive feature of the ML method is that it is statistically well founded, which permits statistical comparison of different models, model parameters and competing trees (Whelan et al. 2001). A disadvantage of the ML method is that it is computationally very demanding. When numbers of sequences become large, it is not possible to examine all possible tree topologies, and a heuristic search for the best tree must then be performed (see below). Although algorithms for heuristic tree searches have greatly been improved in recent years, we can not be sure that the best tree has actually been found (Swofford et al. 1996).

Since the same models of sequence evolution as used for ML methods are also used in

minimum distance methods, one would intuitively think that minimum distance methods are as robust as ML methods in finding the optimal tree. This is however not the case. The most serious criticism of minimum distance is that by the transformation to pairwise genetic distances - each pair of sequences is considered independently of the other sequences - the evolutionary information present at different sites is lost (Steel et al. 1988).

Despite these various pitfalls, it is important to emphasize that if the data contain clear and sound signal for all nodes in a tree, the different phylogenetic reconstruction methods will all converge towards the same, and probably true, phylogenetic tree.

Finding the optimal tree

When the optimality criterion for constructing a phylogenetic tree has been selected, the tree best fitting the criterion has to be found. If the numbers of sequences are small (e.g. <10-14) an exhaustive tree search strategy can be used. An exhaustive tree search strategy is a method which systematically examines all possible relationships between sequences. An exhaustive search is however not possible when the number of sequences becomes too large. This is because the number of trees rises dramatically per newly added sequence. For example, the number of possible trees for nine sequences is 135,135, for ten sequences it is $2x10^6$, for 20 sequences $2x10^{20}$ and for 50 sequences $3x10^{74}$. It is therefore necessary to apply heuristic search strategies, where short-cuts or educated guesses are used for finding the optimal tree.

The most thorough strategy for heuristic tree searches is the branch-and-bound method. The branch-and-bound method will search through the tree space to find the best tree, in such a way that it ignores families of trees that can impossibly be better than the tree already found. Hence, the best tree will be found. Although a branch-and-bound strategy is often faster than an exhaustive search, it is computationally still very intensive when numbers of sequences are large. A heuristic tree search method, which is commonly used, is the Tree Bisection Reconnection (TBR) strategy (Swofford et al. 1996). This method searches for a better tree topology by breaking a tree into two parts, and then reconnecting the two subtrees at all possible branches. If a better tree is found, it will be retained, and another round of TBR is started.

An important first step in heuristic tree searching is the choice of an initial tree for subsequent rearrangements. The closer a starting tree is to the best tree, the lesser number of rearrangements is likely to be necessary before the best tree is found, and the chances of being caught in a local maximum in tree space is the lowest. The initial tree can, for instance, be a randomly selected tree, or one's own favorite best tree, or a neighbor joining tree (see below), or a tree which has been built by stepwise addition of taxa to a growing tree. The stepwise addition method starts with an initial tree of three randomly taken taxa, to which new taxa are subsequently added, one at the time. At each addition the new taxon is evaluated on all branches, and the best tree solution is retained (Swofford et al. 1996).

The most commonly used method to obtain a phylogenetic tree from a minimum distance matrix is the neighbor joining (NJ) method (Saitou and Nei 1987). The NJ method starts with a star-like tree - a tree in which all branches are connected at a single internal node - and successively joins pairs of neighboring taxa such that the total distance in the tree is minimized. After each joining step the two linked taxa are taken as a single 'taxon', and a new tree distance matrix with this 'joined taxon' is calculated before a new round of joining least-distance neighbor taxa is started. An improved

version of the NJ method is the BIONJ method (Gascuel 1997). BioNJ uses the same algorithm as NJ for joining pairs of taxa, but it has a better criterion in selecting pairs of taxa to join, which performs better when distances between taxa are large (Gascuel 1997). A great advantage of the NJ method, to which it thanks its popularity, is the speed by which a tree is constructed. But it should be noted that a NJ-based phylogeny should never be considered as the final best tree based on the minimum distance criterion (Swofford et al. 1996).

Statistical tests of phylogenetic trees

Once the best tree solution has been found, under a certain optimality criterion, one would like to know how reliable this best tree is and to which degree the different relationships in the tree are supported by the data. A commonly used method to assess confidence in an inferred tree is by non-parametric (NP) bootstrapping (usually just called bootstrapping), which first was introduced into phylogenetics by Felsenstein (1985). In NP bootstrapping new data sets are created by random resampling from the original data. The new data sets are of the same size as the original data set, but some sites may be represented more than once, whereas other sites may not be present at all. From the new (bootstrapped) data sets phylogenetic trees are calculated and a consensus tree is made from the different trees. The number of times a node is supported in the consensus tree is taken as the confidence for this node. If a node is supported by 95% or more in the consensus tree it often is taken as significant, whereas support values lower than 75% are considered as non-significant support. But if a node has a support value between 75% and 95% it can be difficult to interpret the significance of such support (Whelan et al. 2001). It has been suggested that NP bootstrapping may be too conservative in evaluating tree confidence (Hillis and Bull 1993). In parsimony, the decay or Bremer support index (Donoghue et al. 1992; Bremer 1994) can be used to judge the support of a monophyletic group (a node in a tree) by calculating the number of extra evolutionary changes needed to make this group paraphyletic (absence of this node in the tree).

Tests for evaluating if there is significant difference between two competing hypotheses are often used in phylogenetic analyses, especially in parsimony and likelihood analyses. What one would like to know is whether an observed difference in score between two trees is significant or could also have happened by chance or random error. For parsimony, the Templeton/winning sites test (Templeton 1983; Prager and Wilson 1988) employs a Wilcoxon ranked sum to calculate the number of characters in favor of each of the two competing trees. A statistical test is then performed to evaluate whether the observed difference in competing characters is significantly different from what could be expected from a random distribution of characters.

Another test, which is commonly used, is the Kishino Hasegawa (KH) test (Kishino and Hasegawa 1989). In the KH test the null hypothesis - that the tree score or length between the two trees is zero - is tested against the score-distribution at each site by a paired t-test, taking the sampling variance into account. If there is no *a priori* expectation that the tree scores of the two trees are different, a two-tailed t-test should be used, whereas if there is an *a priori* expectation that one tree is better than the other, the null hypothesis is violated and the KH test can not be used. Thus, the KH test can only be used for *a priori* competing hypotheses, such as derived from independent datasets or other evidence, and not for *a posteriori* competing hypothesis such as comparing the best tree with the second best tree solution (see Swofford et al. 1996; Goldman et al. 2000).

One way to test if there is significant difference between two *a posteriori* competing hypotheses is by performing parametric bootstrapping (also known as Monte Carlo simulation) (Hillis et al. 1996; Hulsenbeck et al. 1996; Shimodaira and Hasegawa 1999; Goldman et al. 2000). In parametric bootstrapping new data sets with the same length are simulated, based e.g. on the null-hypothesis (such as the second best tree) and the same model of sequence evolution used to find the two competing hypotheses. The newly simulated data sets can then be used to estimate new tree scores of the null-hypothesis (the second best tree). The distribution of the differences in tree scores between the simulated data and the null-hypothesis as compared to the difference between the two competing hypothesis (the best and second best tree solution) is then used to judge if there is significant difference between the two competing hypothesis. Besides evaluating the significance of differences between *a posteriori* competing hypothesis, parametric bootstrapping is also very useful for various other purposes, such as determining the sources of systematic bias (e.g. long branch attraction), sequence model fitting and predicting how much data may be needed for getting a reliable phylogeny (Hillis et al. 1996).

Genes used in this thesis

Alpha 2B adrenergic receptor

The alpha 2B adrenergic receptor (A2AB) belongs to the family of G protein-coupled receptors (GPCRs). The GPCR family is with more than 616 members in the human genome (Venter et al. 2001) the largest family of cell-surface molecules involved in signal transmission. A variety of ligands can activate GPCRs, like hormones, odorant molecules, peptides and neurotransmitters. As the name indicates, GPCRs can interact with heterotrimeric G proteins (α, ß and γ subunits) and upon activation by a ligand a conformational change allows exchange of GDP for GTP on the G protein α-subunit. The activated G_{α} and $G_{\beta/\gamma}$ subunits then stimulate effector molecules such as adenylyl cyclase and ion channels, which subsequently leads to the activation or inhibition of the production of secondary messengers like cAMP. Besides this 'traditional' signaling pathway, research has indicated in recent years that activation of GPCRs can stimulate a variety of other signaling pathways, of which several are not mediated by G proteins (Marinissen and Gutkind 2001; Perry and Lefkowitz 2002).

The regulation of GPCR activity is mediated by arrestin-dependent desensitization and internalization. When a ligand binds to the receptor, the latter becomes phosphorylated. Arrestins can subsequently bind to the liganded receptor, which leads to uncoupling of G proteins from the receptor, and to receptor internalization into endosomes. In the endosomes the GPCR can either be recycled back to the cell-surface or trafficked to lysosomes for degradation (Pierce and Lefkowitz 2001; Tsao et al. 2001). In the cell membrane GPCRs form dimeric or oligomeric complexes, which can be homo- or heteromeric. The state of these complexes (homo-, hetero-, di- or oligomeric) has been shown to be an important determinant in regulating ligand binding and thereby signaling (Devi 2001).

All GPCRs are characterized by a structure comprising a bundle of seven transmembrane (TM) helical domains, connected by three extra- and three intracellular loops, and have an extracellular N-terminus and an intracellular C-terminus (figure 4; Baldwin 1993). The GPCR family

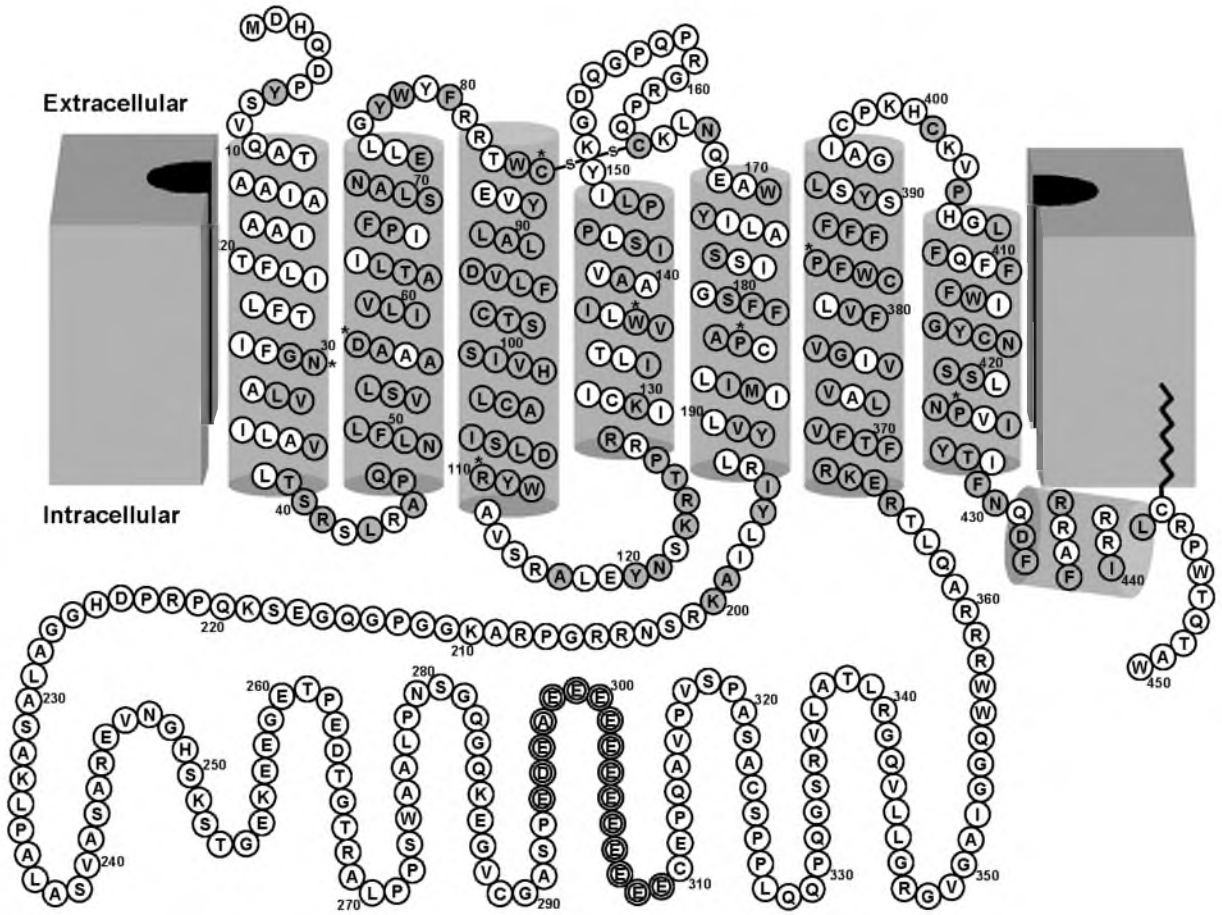


Figure 4. Predicted secondary structure of the human alpha 2B adrenergic receptor (A2AB) based on the crystal structure of bovine rhodopsin (Palczewski et al. 2000). The gray cylinders indicate that in addition to the seven transmembrane (TM) helices there is a short α-helix in the intracellular C-terminal tail, which is attached to the membrane through a palmitoylated cysteine, as indicated. Residues conserved in the alpha 2 adrenergic receptor family (A2AA, A2AB, A2AC & A2AR; SWISS-PROT names) are marked in gray. Double lined circles indicate the 'glutamic acid' domain in the very long third intracellular loop. Conserved residues used to align A2AB with bovine rhodopsin, to enable the prediction of the TM α-helical regions, are indicated by asterisks (*).

can be divided into six groups: class A to E and the frizzled/smoothened family (Bockaert and Pin 1999; Horn et al. 2001; http://www.gpcr.org/7tm/). Class A receptors, to which A2AB belongs, form by far the largest and best studied family of GPCRs, and are characterized by binding mostly small ligands inside the helical bundel. They have a conserved aspartate in the second TM helix, a tripeptide DRY/EWY at the intracellular end of the third TM helix, which is involved in G protein coupling, several other conserved residues in their TM domains, and a cysteine in the C-terminal tail which is palmitoylated (Ceresa and Limbird 1994; Oliviera et al. 1994; Strader et al. 1995; Scheer et al. 1996). The crystal structure is only known for one GPCR, namely rhodopsin, which is a class A receptor (Palczewski et al. 2000).

A2AB is a receptor for catecholamines, such as adrenaline. Together with the two other alpha 2 adrenergic receptors A and C it plays an important role in various cellular processes such as central and peripheral control of circulation and control of secretion. Knockout studies in mouse have shown that lack of A2AB influences viability, response to salt-induced hypertension, and blood pressure changes upon agonist treatment (Link et al. 1996, Makaritsis et al. 1999 & 2000). A2AB mainly mediates the inhibition of adenylyl cyclase via coupling to $G_{i/o}$ proteins (Limbird 1988), but

stimulation of adenylyl cyclase via coupling to the G_s proteins has also been reported (Eason et al. 1992). This biphasic response seems to depend on the concentration of agonist, where low concentrations lead to inhibition and high concentration to stimulation of adenylyl cyclase.

A2AB has a length of about 450 amino acid residues (figure 4), and is encoded by a single copy intronless gene located on chromosome 2 in human (Lomasney et al. 1990). The expression of A2AB in different tissues has not been totally clarified, but a distinct expression in various tissues, mostly in the periphery and with highest expression in the kidney has been reported (e.g., Eason and Liggett 1993; Link et al. 1996). Special features of A2AB are the lack of N-linked glycosylation sites in the extracellular domains, and a very long third intracellular loop which contains a characteristic glutamic acid repeat (figure 4). This glutamic acid domain has been suggested to be required for agonist-promoted phosphorylation of the third intracellular loop and desensitization of A2AB, since deletion of this repeat or replacement by glutamines abolishes these effects of agonist treatment (Jewell-Motz and Liggett 1995). The importance of this acidic strand was further confirmed by the phenotype of an allelic deletion mutant in human. This mutant, which lacks three glutamic acid residues in the acidic strand of the third intracellular loop, shows a significant decrease in agonist-promoted phosphorylation and lack of desensitization (Small et al. 2001). The mutant allele has also been associated with reduced metabolic rate in obese subjects, and is a genetic risk factor for acute coronary events (Heinonen et al. 1999; Snapir et al. 2001), but does not play a role in hypertension (Baldwin et al. 1999). The third intracellular loop has also been shown to be necessary for microtubule-dependent sorting of A2AB to the cell surface (Saunders and Limbird 1999).

Aquaporin 2

Aquaporin 2 (AQP2) is a member of the aquaporin (AQP) family, formerly called the major intrinsic protein (MIP) family, which are channel-proteins that selectively facilitate transport of small molecules such as water, urea and glycerol across cell membranes. The AQP family can be classified into two major clusters, the AQP cluster and the glycerol facilitator-like protein (GLP) cluster (Heymann and Engel 1999; Zardoya and Villalba 2001). The two clusters are suggested to be descendents of two homologous bacterial genes with similar functions (Park and Saier 1996). Ten representatives of the AQP family are present in mammals, with seven (AQP 0, 1, 2, 4, 5, 6, 8) being selective in transport of water and belonging to the AQP cluster, and three (AQP 3, 7 and 9) being permeable for water, glycerol and urea, and belonging to the GLP cluster (Heymann and Engel 1999; Zardoya and Villalba 2001). All AQPs are thought to be structurally more or less similar, and recently the crystal structure of AQP1 has been determined (Murata et al. 2000; Ren et al. 2001). The structure of AQPs consist of six tilted transmembrane (TM) α-helices (figure 5) which form a right-handed bundle. The six TM helices are connected by five loops (A-E in figure 5), with the N- and C-termini located intracellularly (e.g. Preston et al. 1994; Murata et al. 2000). AQP sequences consist of two repeated halves, each comprising three TM helices, and supposed to be the result of an intragenic duplication (Wistow et al. 1991; Reizer et al. 1993). In loop B and E a highly conserved motif Asn-Pro-Ala is present; these two motifs dip into the membrane from opposite sides, thereby forming the aqueous pore as was predicted by the hourglass model (figure 5, bottom left) (Jung et al. 1994; Murata et al. 2000). Most AQPs form homotetramers in the cell membrane (Verbavatz et al. 1993; van Hoek et al. 1998), with each monomer being a functional unit (Preston et al. 1993; Shi et al. 1994).

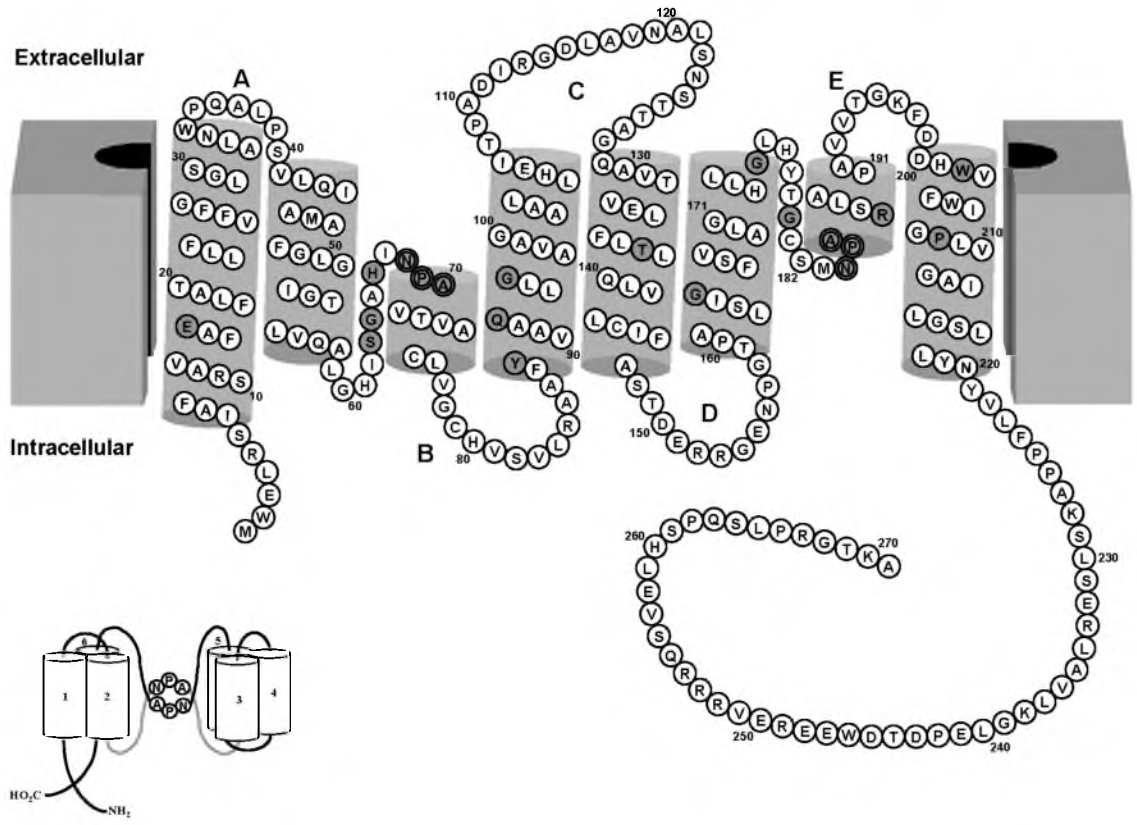


Figure 5. Secondary structure of human aquaporin (AQP) 2 as predicted from the crystal structure of human AQP 1 (Murata et al. 2000). Residues marked gray are conserved among the seven human AQPs belonging to the AQP cluster (see text). Double lined residues indicate the highly conserved Asn-Pro-Ala motifs in loops B and E, which bend back into the membrane to form the aqueous pore as predicted by the hourglass model (bottom left).

AQP2 has a length of 271 amino acid residues. It is encoded by a single copy gene, comprised of four exons, and located in an AQP gene cluster on chromosome 12 in human (Uchida et al. 1994; Deen et al. 1994a; Ma et al. 1996). AQP2 is exclusively expressed in the kidney collecting duct where it is located in the apical plasma membrane and apical vesicles (Fushimi et al. 1993; Nielsen et al. 1993; Ma et al. 1994). It plays an exclusive role in the vasopressin-dependent concentration of urine (Deen et al. 1994b). Mutations in the human AQP2 gene have been shown to cause nephrogenic diabetes insipidus (e.g., Deen et al. 1994b; Tamarappoo and Verkman 1998; Kuwahara et al. 2001). Patients with this disease are unable to produce concentrated urine, which consequently can lead to dehydration. One of these mutants has been shown to result in retention of AQP2 in the Golgi complex (Mulder et al. 1998), whereas all other AQP2 mutants are impaired in their export from the endoplasmic reticulum, probably because of misfolding of the mutated AQP2 proteins (Deen et al. 1995; Mulder et al. 1997).

Aim and outline of thesis

The aim of this thesis project was to resolve the phylogenetic relationships amongst mammals in general, and eutherians in particular, by means of molecular sequence comparisons. Simultaneously, the molecular evolution of the nuclear-encoded genes to be compared should be described. The importance of achieving a robust and well resolved tree, which consequently - or hopefully - will be the true mammalian tree, are multifarious: 1) in relation to biomedical studies it can help to decide which animals are the best choice as model animals for human diseases; 2) in genome research it can help choosing mammalian genomes for sequencing; 3) it is important to determine if proteins or species display deviating molecular evolutionary mechanisms; 4) it can be used as a scaffold for reconstructing morphological evolution; and 5) it can be related to biogeographic knowledge.

In chapter 2 the molecular evolution of exon 1 of aquaporin 2 is analyzed, and further support is given that aardvarks and elephant shrews group together with Paenungulata (elephants, sea cows and hyraxes) in an African clade (later dubbed Afrotheria).

In the following chapters (3-5) Afrotheria is recognized to also include two families of endemic African insectivores, golden moles and tenrecs, consequently necessitating the division of the order Insectivora into two groups, one group with African origin (Afrosoricida) and one with Laurasian origin (Eulipotyphla). Furthermore, the radiation of the different Afrotherian orders, which all have African origins, is related to Africa's isolation during the mid Cretaceous to early Cenozoic.

The division of the eutherian orders in four major groups is described, for the first time, in chapter 6. This result is achieved by combining different genes into a single data matrix, thereby getting better resolution at the deep level of the eutherian tree. This four-partition of eutherians is subsequently used to describe parallel adaptive radiations in two of the four major clades of eutherian mammals.

Chapter 7 describes the molecular evolution of the mammalian alpha 2B adrenergic receptor, determined in the context of the previous phylogenetic studies, in relation to its structure and function.

Finally, in chapter 8, the most important results and conclusions of this study are summarized and discussed in relation to recent developments in the field, and with regard to their implications for further research.

References:

Alvarez LW, Alvarez W, Asaro F, and Michel HV. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science* **208**: 1095-1108

Adachi J, and Hasegawa M. 1995. MOLPHY: Program for molecular phylogenetics. Ver. 2.3, Institute of Statistical Mathematics, Japan

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, and Janke A. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc. Natl. Acad. Sci. USA* **99**: 8151-8156

Baldwin JM. 1993. The probable arrangement of the helices in G protein-coupled receptors. *EMBO J.* **12**: 1693-1703

Baldwin CT, Schwartz F, Baima J, et al. (14 co-authors). 1999. Identification of a polymorphic glutamic acid stretch in the alpha2B-adrenergic receptor and lack of linkage with essential hypertension. *Am. J. Hypertens.* **12**: 853-857

Bockaert J, and Pin JP. 1999. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.* **18**: 1723-1729

Bremer K. 1994. Branch support and tree stability. *Cladistics* **10**: 295-304

Bromham L, Phillips MJ, and Penny D. 1999. Growing up with dinosaurs: molecular dates and the mammalian radiation. *Trends Ecol. Evol.* **14**: 113-118

Cavalli-Sforza LL, and Edwards AW. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* **32**: 550-570

Cao Y, Fujiwara M, Nikaido M, Okada N, and Hasegawa M. 2000 Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. *Gene* **259**: 149-158

Chang JT. 1996. Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. *Math. Biosci.* **137**: 51-73

Ceresa BP, and Limbird LE. 1994. Mutation of an aspartate residue highly conserved among G-protein-coupled receptors results in nonreciprocal disruption of alpha 2-adrenergic receptor-G-protein interactions. A negative charge at amino acid residue 79 forecasts alpha 2A-adrenergic receptor sensitivity to allosteric modulation by monovalent cations and fully effective receptor/G-protein coupling. *J. Biol. Chem.* **269**: 29557-29564

Dayhoff MO, Schwartz RM, and Orcutt BC. 1978. A model of evolutionary change in proteins. pp. 345-352 in Atlas of Protein Sequence and Structure, volume 5, supplement 3, ed. M. O. Dayhoff. National Biomedical Research Foundation, Silver Spring, Maryland .

de Jong WW. 1998. Molecules remodel the mammalian tree. *Trends Ecol. Evol.* **13**: 270-275

Deen PM, Weghuis DO, Sinke RJ, Geurts van Kessel A, Wieringa B, and van Os CH. 1994a. Assignment of the human gene for the water channel of renal collecting duct Aquaporin 2 (AQP2) to chromosome 12 region q12-->q13. *Cytogenet. Cell Genet.* **66**: 260-262

Deen PM, Verdijk MA, Knoers NV, Wieringa B, Monnens LA, van Os CH, and van Oost BA. 1994b. Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. *Science* **264**: 92-95

Deen PM, Croes H, van Aubel RA, Ginsel LA, and van Os CH. 1995. Water channels encoded by mutant aquaporin-2 genes in nephrogenic diabetes insipidus are impaired in their cellular routing. *J. Clin. Invest.* **95**: 2291-2296

D'Erchia AM, Gissi C, Pesole G, Saccone C, and Arnason U. 1996. The guinea-pig is not a rodent. *Nature* **381**: 597-600

Devi LA. 2001. Heterodimerization of G-protein-coupled receptors: pharmacology, signaling and trafficking. *Trends Pharmacol. Sci.* **22**:532-537

Donoghue MJ, Olmstead TG, Smith JF, and Palmer JD. 1992. Phylogenetic relationships of Dipsacales based on rbcL sequences. *Ann. Missouri Bot. Garden* **79**: 333-345

Eason MG, Kurose H, Holt BD, Raymond JR, and Liggett SB. 1992. Simultaneous coupling of alpha 2-adrenergic receptors to two G proteins with opposing effects. Subtype-selective coupling of alpha 2C10, alpha 2C4, and alpha 2C2 adrenergic receptors to Gi and Gs. *J. Biol. Chem.* **267**: 15795-15801

Eason MG, and Liggett SB. 1993. Human alpha 2-adrenergic receptor subtype distribution: widespread and subtype-selective expression of alpha 2C10, alpha 2C4, and alpha 2C2 mRNA in multiple tissues. *Mol. Pharmacol.* **44** :70-75

Easteal S. 1999. Molecular evidence for the early divergence of placental mammals. *Bioessays* **21**: 1052-1058

Eizirik E, Murphy WJ, and O'Brien SJ. 2001. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* **92**: 212-219

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**: 401-410

Felsenstein J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17**: 368-376

Felsenstein J. 1984. Distance methods for inferring phylogenies: a justification. *Evolution* **38**: 16-24

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. **39**: 783-791

Felsenstein J. 2001. PHYLIP (Phylogeny Inference Package) version 3.6a2.1. (Department of Genetics, University of Washington, Seattle)

Felsenstein J. 2001. The troubled growth of statistical phylogenetics. *Syst. Biol.* **50**: 465-467

Fitch WM. 1971. Toward defining

the course of evolution: minimum change for a specified tree topology. *Syst. Zool.* **20**: 406-416.

Foote M, Hunter JP, Janis CM, and Sepkoski JJ Jr. 1999. Evolutionary and preservational constraints on origins of biologic groups: Divergence times of eutherian mammals. *Science* **283**: 1310-1314

Fushimi K, Uchida S, Hara Y, Hirata Y, Marumo F, and Sasaki S. 1993. Cloning and expression of apical membrane water channel of rat kidney collecting tubule. *Nature* **361**: 549-552

Gatesy J, and O'Leary MA. 2001. Deciphering whale origins with molecules and fossils. *Trends Ecol. Evol.* **16**: 562-570

Gascuel O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**: 685-695

Goldman N. 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and parsimony analysis. *Syst. Zool.* **39**: 345-361

Goldman N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36**: 182-198

Goldman N, Anderson JP, and Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**: 652-670

Hasegawa M, Kishino H, and Yano T. 1985a. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**: 160-174

Hasegawa M, Iida Y, Yano T, Takaiwa F, and Iwabuchi M. 1985b. Phylogenetic relationships among eukaryotic kingdoms inferred from ribosomal RNA sequences. *J. Mol. Evol.* **22**: 32-38

Hillis DM, and Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**: 182-192

Hillis DM, Mable BK, and Moritz C. 1996. Applications of molecular systematics: The state of the field and a look to the future. In: Molecular Systematics (second edition), Hillis DM, Moritz C, and Mable BK. eds., pp. 515-543, Sinauer Associates, Sunderland, Massachusetts.

Hedges SB, Parker PH, Sibley CG, and Kumar S. 1996. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**: 226-229

Hedges SB, and Kumar S. 1999. Divergence times of eutherian mammals. *Science* **285**: Technical comments 2031a

Heinonen P, Koulu M, Pesonen U, Karvonen MK, Rissanen A, Laakso M, Valve R, Uusitupa M, and Scheinin M. 1999. Identification of a three-amino acid deletion in the alpha2B-adrenergic receptor that is associated with reduced basal metabolic rate in obese subjects. *J. Clin. Endocrinol. Metab.* **84**: 2429-2433

Heymann JB, and Engel A. 1999. Aquaporins: Phylogeny, Structure, and Physiology of Water Channels. *News Physiol. Sci.* **14**: 187-193

Horn F, Vriend G, and Cohen FE. 2001. Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems. *Nucleic. Acids Res.* **29**: 346-349

Huelsenbeck JP, Hillis DM, and Nielsen R.1996. A likelihood-ratio test of monophyly. *Syst. Biol.* **45**: 546-558

Huelsenbeck JP, and Rannala B. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* **276**: 227-232

Huelsenbeck JP. 1997. Is the Felsenstein zone a fly trap? *Syst. Biol.* **46**: 69-74

Janke A, Magnell O, Wieczorek G, Westerman M, and Arnason U. 2002. Phylogenetic Analysis of 18S rRNA and the Mitochondrial Genomes of the Wombat, Vombatus ursinus, and the Spiny Anteater, Tachyglossus aculeatus: Increased Support for the Marsupionta Hypothesis. *J. Mol. Evol.* **54**: 71-80

Jewell-Motz EA, and Liggett SB. 1995. An acidic motif within the third intracellular loop of the alpha2C2 adrenergic receptor is required for agonist-promoted phosphorylation and desensitization. *Biochemistry* **34**: 11946-11953

Jones DT, Taylor WR, and Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Comp. Appl. Biosci.* **8**: 275-282

Jukes TH, and Cantor CR. 1969. Evolution of protein molecules. pp. 21-132 in Mammalian Protein Metabolism, ed. H. N. Munro. Academic Press, New York

Jung JS, Preston GM, Smith BL, Guggino WB, and Agre P. 1994. Molecular structure of the water channel through aquaporin CHIP. The hourglass model. *J. Biol. Chem.* **269**: 14648-14654

Killian JK, Buckley TR, Stewart N, Munday BL, and Jirtle RL. 2001. Marsupials and Eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution. *Mamm. Genome* **12**: 513-517

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111-120

Kimura M. 1981. Possibility of extensive neutral evolution under stabilizing selection with special reference to nonrandom usage of synonymous codons. *Proc. Natl. Acad. Sci. USA* **78**: 5773-5777

Kishino H, and Hasegawa M. 1989. Evaluation of the maximum likelihood estimates of the evolutionary tree topologies from sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**: 170-179

Kishino H, Miyata T, and Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **30**:151-160

Kleineidam RG, Pesole G, Breukelman HJ, Beintema JJ, and Kastelein RA. 1999. Inclusion of cetaceans within the order Artio-

dactyla based on phylogenetic analysis of pancreatic ribonuclease genes. *J. Mol. Evol.* **48**:360-368

Krettek A, Gullberg A, and Arnason U. 1995. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, Erinaceus europaeus, and the phylogenetic position of the Lipotyphla. *J. Mol. Evol.* **41**: 952-957

Kumar S, and Hedges SB. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**: 917-920

Kuwahara M, Iwai K, Ooeda T, Igarashi T, et al. (11 co-authors). Three families with autosomal dominant nephrogenic diabetes insipidus caused by aquaporin-2 mutations in the C-terminus. *Am. J. Hum. Genet.* **69**: 738-748

Ladoukakis ED, and Zouros E. 2001. Direct evidence for homologous recombination in mussel (Mytilus galloprovincialis) mitochondrial DNA. *Mol. Biol. Evol.***18**: 1168-1175

Lake JA. 1994. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc. Natl. Acad. Sci. USA* **91**: 1455-1459

Lanave C, Preparata G, Saccone C, and Serio G. 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20**: 86-93

Limbird LE. 1988. Receptors linked to inhibition of adenylate cyclase: additional signaling mechanisms. *FASEB J.* **2**: 2686-2695

Lin YH, and Penny D. 2001. Implications for bat evolution from two new complete mitochondrial genomes. *Mol. Biol. Evol.* **18**: 684-688

Link, RE, Desai K, Hein L, Stevens ME, Chruscinski A, Bernstein D, Barsh GS, Kobilka BK. 1996. Cardiovascular regulation in mice lacking alpha2-adrenergic receptor subtypes b and c. *Science* **273**: 803-805

Lockhart, PJ, Steel MA, Hendy MD, and Penny D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**: 605-612

Lomasney JW, Lorenz W, Allen LF, King K, Regan JW, Yang-Feng TL, Caron MG, and Lefkowitz RJ. 1990. Expansion of the alpha 2-adrenergic receptor family: cloning and characterization of a human alpha 2-adrenergic receptor subtype, the gene for which is located on chromosome 2. *Proc. Natl. Acad. Sci. USA* **87**: 5094-5098

Ma T, Hasegawa H, Skach WR, Frigeri A, and Verkman AS. 1994. Expression, functional analysis, and in situ hybridization of a cloned rat kidney collecting duct water channel. *Am. J. Physiol.* **266**: C189-197

Ma T, Yang B, Kuo WL, and Verkman AS. 1996. cDNA cloning and gene structure of a novel water channel expressed exclusively in human kidney: evidence for a gene cluster of aquaporins at chromosome locus 12q13. *Genomics* **35**: 543-550

Macdonald DW. 1984. The Encyclopedia of Mammals. Andromeda Oxford Ltd: Oxford

Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM, Stanhope MJ, deJong WW, and Springer MS. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* **409**: 610–614.

Makaritsis KP, Handy DE, Johns C, Kobilka B, Gavras I, and Gavras H. 1999. Role of the alpha2B-adrenergic receptor in the development of salt-induced hypertension. *Hypertension* **33**: 14-17

Makaritsis KP, Johns C, Gavras I, and Gavras H. 2000. Role of alpha(2)-adrenergic receptor subtypes in the acute hypertensive response to hypertonic saline infusion in anephric mice. *Hypertension* **35**: 609-613

Marinissen MJ, and Gutkind JS. 2001. G protein-coupled receptors and signaling networks: emerging paradigms. *Trends Pharmacol. Sci.* **22**: 368-376

McKenna, MC. and Bell, SK. 1997. Classification of Mammals Above the Species Level, Columbia Univ. Press, New York

Montgelard C, Catzeflis FM, and Douzery E. 1997. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome b and 12S rRNA mitochondrial sequences. *Mol. Biol. Evol.* **14**: 550-559

Mouchaty SK, Gullberg A, Janke A, and Arnason U. 2000. Phylogenetic position of the tenrecs (Mammalia: Tenrecidae) of Madagascar based on analysis of the complete mitochondrial genome sequence of Echinops telfairi. *Zool. Scripta* **29**: 307-317

Mulders SM, Knoers NV, Van Lieburg AF, Monnens LA, Leumann E, Wuhl E, Schober E, Rijss JP, Van Os CH, and Deen PM. 1997. New mutations in the AQP2 gene in nephrogenic diabetes insipidus resulting in functional but misrouted water channels. *J. Am. Soc. Nephrol.* **8**: 242-248

Mulders SM, Bichet DG, Rijss JP, Kamsteeg EJ, Arthus MF, Lonergan M, Fujiwara M, Morgan K, Leijendekker R, van der Sluijs P, van Os CH, and Deen PM. 1998. An aquaporin-2 water channel mutant which causes autosomal dominant nephrogenic diabetes insipidus is retained in the Golgi complex. *J. Clin. Invest.* **102**: 57-66

Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann JB, Engel A, and Fujiyoshi Y. 2000. Structural determinants of water permeation through aquaporin-1. *Nature* **407**: 599-605

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, and O'Brien SJ. 2001a. Molecular phylogenetics and the origins of placental mammals. *Nature* **409**: 614–618

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, and Springer MS. 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**: 2348-2351

Nielsen S, DiGiovanni SR, Christensen EI, Knepper MA, and Harris HW. 1993. Cellular and subcellular immunolocalization of vasopressin-regulated water channel in rat kidney. *Proc. Natl. Acad. Sci. USA* **90**: 11663-11667

Nikaido M, Rooney AP, and Okada N. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA* **96**: 10261-10266

Nikaido M, Kawai K, Cao Y, Harada M, Tomita S, Okada N, and Hasegawa M. 2001. Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* **53**: 508-516

Novacek, MJ. 1992. Mammalian phylogeny: Shaking the tree. *Nature* **356**: 121-125

Oliveira L, Paiva AC, Sander C, and Vriend G. 1994. A common step for signal transduction in G protein-coupled receptors. *Trends Pharmacol. Sci.* **15**: 170-172

Palczewski K, Kumasaka T, Hori T, et al., (12 co-authors). 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**: 739-745

Park JH, and Saier MH Jr. 1996. Phylogenetic characterization of the MIP family of transmembrane channel proteins. *J. Membr. Biol.* **153**: 171-180

Penny D, Hasegawa M, Waddell PJ, and Hendy MD. 1999. Mammalian evolution: Timing and implications from using the LogDeterminant transform for proteins of differing amino acid composition. *Syst. Biol.* **48**: 76-93

Perry SJ, and Lefkowitz RJ. 2002. Arresting developments in heptahelical receptor signaling and regulation. *Trends Cell Biol.* **12**: 130-138.

Pierce KL, and Lefkowitz RJ. 2001. Classical and new roles of beta-arrestins in the regulation of G-protein-coupled receptors. *Nat. Rev. Neurosci.* **2**: 727-733

Prager E, and Wilson AC. 1988. Ancient origin of lactalbumin from lysozyme: Analysis of DNA and amino acid sequences. *J. Mol. Evol.* **27**: 326-335.

Preston GM, Jung JS, Guggino WB, and Agre P. 1993. The mercury-sensitive residue at cysteine 189 in the CHIP28 water channel. *J. Biol. Chem.* **268**: 17-20

Preston GM, Jung JS, Guggino WB, and Agre P. 1994. Membrane topology of aquaporin CHIP. *J. Biol. Chem.* **269**: 1668-1673

Pumo DE, Finamore PS, Franek WR, Phillips CJ, Tarzami S, and Balzarano D. 1998. Complete mitochondrial genome of a neotropical fruit bat, Artibeus jamaicensis, and a new hypothesis of the relationships of bats to other eutherian mammals. *J. Mol. Evol.* **47**: 709-717

Reizer J, Reizer A, and Saier MH Jr. 1993. The MIP family of integral membrane channel proteins: sequence comparisons, evolutionary relationships, reconstructed pathway of evolution, and proposed functional differentiation of the two repeated halves of the proteins. *Crit. Rev. Biochem. Mol. Biol.* **28**: 235-257

Ren G, Reddy VS, Cheng A, Melnyk P, and Mitra AK. 2001. Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc. Natl. Acad. Sci. USA* **98**: 1398-1403

Reyes A, Pesole G, and Saccone C. 2000. Long-branch attraction phenomenon and the impact of among-site rate variation on rodent phylogeny. *Gene* **259**: 177-187

Rodriguez F, Oliver JL, Marin A, and Medina JR. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* **142**: 485-501

Rogers JS. 1997. On the consistency of maximum likelihood estimation of phylogenetic trees from nucleotide sequences. *Syst. Biol.* **46**:354-357

Saitou N, and Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406-425

Saunders C, and Limbird LE. 2000. Microtubule-dependent regulation of alpha(2B) adrenergic receptors in polarized MDCKII cells requires the third intracellular loop but not G protein coupling. *Mol. Pharmacol.* **57**: 44-52

Scheer A, Fanelli F, Costa T, De Benedetti PG, and Cotecchia S. 1996. Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. *EMBO J.* **15**: 3566-3578

Schmitz J, Ohme M, and Zischler H. 2000. The complete mitochondrial genome of Tupaia belangeri and the phylogenetic affiliation of scandentia to other eutherian orders. *Mol. Biol. Evol.* **17**: 1334-1343

Shi LB, Skach WR, and Verkman AS. 1994. Functional independence of monomeric CHIP28 water channels revealed by expression of wild-type mutant heterodimers. *J. Biol. Chem.* **269**: 10417-10422

Shimodaira H, and Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**: 1114-1116

Shoshani J, and McKenna MC. 1998. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. *Mol Phylogenet Evol.* **9**: 572-584

Small KM, Brown KM, Forbes SL, and Liggett SB. 2001. Polymorphic deletion of three intracellular acidic residues of the alpha 2B-adrenergic receptor decreases G protein-coupled receptor kinase-mediated phosphorylation and desensitization. *J. Biol. Chem.* **267**: 4917-4922

Smith AG, Smith DG, and Funnell BM. 1994. Atlas of Mesozoic and Cenozoic coastlines. Cambridge University Press: Cambridge

Snapir A, Heinonen P, Tuomainen TP, et al. (14 co-authors). 2001. An insertion/deletion polymorphism in the alpha2B-adrenergic receptor gene is a novel genetic risk factor for acute coronary events. *J. Am. Coll. Cardiol.* **37**: 1516-1522

Springer MS, Cleven GC, Madsen O, de Jong WW, Waddell VG, Amrine HM, and Stanhope MJ. 1997. Endemic African mammals shake the phylogenetic tree.

Nature **388**: 61-64

Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, de Jong WW, and Stanhope MJ 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* **18**: 132-143

Springer MS, and de Jong WW. 2001. Which mammalian supertree to bark up? *Science* **291**: 1709-1711

Stanhope MJ, Waddell VG, Madsen O, de Jong WW, Hedges SB, Cleven GC, Kao D, and Springer MS. 1998. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl. Acad. Sci. USA* **95**: 9967-9972

Steel MA, Hendy MD, and Penny D. 1988. Loss of information in genetic distances. *Nature* **336**: 118

Steel MA. 1994. Recovering a tree from the Markov leaf colourations it generates under a Markov model. *Appl. Math. Lett.* **7**: 19-23

Steel MA, and Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Mol. Biol. Evol.* **17**: 839-850

Strader CD, Fong TM, Graziano MP, and Tota MR. 1995. The family of G protein-coupled receptors. *FASEB J.* **9**: 745-754

Swofford DL, Olsen GP, Waddell PJ, and Hillis DM. 1996. Phylogenetic inference. In: Molecular Systematics, D. M. Hillis, C. Moritz, and B. K. Mable, eds., pp. 407-492, Sinauer, Massachusetts

Swofford DL. 1998. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tajima F, and Nei M. 1984. Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* **1**: 269-285

Tamarappoo BK, and Verkman AS. 1998. Defective aquaporin-2 trafficking in nephrogenic diabetes insipidus and correction by chemical chaperones. *J. Clin. Invest.* **101**: 2257-2267

Tamura K, and Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**: 512-526

Tsao P, Cao T, and von Zastrow M. 2001. Role of endocytosis in mediating downregulation of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **22**: 91-96

Templeton AR. 1983. Phylogenetic inference from re-striction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37**: 221-244.

Uchida S, Sasaki S, Fushimi K, and Marumo F. 1994. Isolation of human aquaporin-CD gene. *J. Biol. Chem.* **269**: 23451-23455

Ursing BM and Arnason U. 1998. Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade. *Proc. R. Soc. Lond. B. Biol. Sci.* **265**: 2251-2255

van Hoek AN, Yang B, Kirmiz S, and Brown D. 1998. Freeze-fracture analysis of plasma membranes of CHO cells stably expressing aquaporins 1-5. *J. Membr. Biol.* **165**: 243-254

Venter JC, Adams MD, Myers EW, et al., (274 co-authors). 2001. The sequence of the human genome. (published erratum appears in Science (2001) 292:1838) *Science* **291**: 1304-1351

Verbavatz JM, Brown D, Sabolic I, Valenti G, Ausiello DA, Van Hoek AN, Ma T, and Verkman AS. 1993. Tetrameric assembly of CHIP28 water channels in liposomes and cell membranes: a freeze-fracture study. *J. Cell Biol.* **123**: 605-618

Whelan S, Liò P, and Goldman N. 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Ecol. Evol.* **17**: 262-272

Wistow GJ, Pisano MM, and Chepelinsky AB. 1991. Tandem sequence repeats in transmembrane channel proteins. *Trends Biochem. Sci.* **16**: 170-171

Yang Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* **10**: 1396-1401

Yang Z, Goldman N, and Friday A. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol. Biol. Evol.* **11**:316-324

Zardoya R, and Villalba S. 2001. A phylogenetic framework for the aquaporin family in eukaryotes. *J. Mol. Evol.* **52**: 391-404

Chapter Two

Molecular Evolution of Mammalian Aquaporin-2: Further Evidence that Elephant Shrew and Aardvark Join the Paenungulate Clade

Ole Madsen, Peter M.T. Deen, Graziano Pesole, Cecilia Saccone and Wilfried W. de Jong. 1997. Molecular Evolution of Mammalian Aquaporin-2: Further Evidence that Elephant Shrew and Aardvark Join the Paenungulate Clade. *Mol. Biol. Evol.* 14: 363-371.

A 328-bp sequence from exon 1 of the gene for aquaporin-2 (AQP2) was compared in 12 mammalian species, representing as many eutherian orders. This sequence encodes the N-terminal half of this kidney-specific water channel protein. Most amino acid replacements, as well as an insertion, have occurred in extracellular loops connecting the transmembrane helices, in agreement with a lower functional importance of these loops. Phylogenetic analyses were performed with parsimony, distance and maximum likelihood methods. The AQP2 data set, alone as well as in combination with previously published αA-crystallin protein sequences, strongly supports a clade consisting of elephant, hyrax, aardvark and elephant shrew, reaching bootstrap values of 99%. This finding fully agrees with the only other presently available sequence data sets that include these taxa, those of von Willebrand factor and interphotoreceptor-retinoid binding protein, and suggests that this extended paenungulate clade is one of the most conspicuous superordinal groupings in eutherian phylogeny. Some support was obtained for an artiodactyl/perissodactyl clade, while the grouping of pholidotes with edentates was contradicted.

Introduction

The divergence of the major groups of placental mammals occurred in the Cretaceous, about 100 million years ago (Archibald 1996; Hedges et al. 1996). Because of their relatively rapid radiation, the resolution of the phylogenetic relationships between the 18 orders of living eutherian mammals is intrinsically difficult. Indeed, comparison of recent reviews about eutherian ordinal phylogeny reveals distressingly little consensus, both within and between molecules and morphology (Benton 1988; Novacek 1992; Graur 1993; Honeycutt and Adkins 1993; Szalay Novacek, and McKenna 1993; Allard, McNiff, and Miyamoto 1996). A major reason for the as yet limited progress in the resolution of mammalian ordinal phylogeny is the lack of adequate molecular data sets, covering orthologous sequences from a sufficiently broad array of orders. This is especially true for such traditionally problematic orders as Pholidota (pangolins), Hyracoidea (hyraxes), Tubulidentata (aardvarks) and Macroscelidea (elephant shrews). Pholidota have usually been considered a sistergroup to the Edentata (sloths, anteaters and armadillos) (Patterson 1978; Novacek 1992), but certain morphological studies (McKenna 1975; Rose and Emry 1993) and the scanty protein sequence and serological data (de Jong and Goodman 1982; Shoshani 1986; Czelusniak et al. 1990; de Jong, Leunissen, and Wistow 1993; Sarich 1993) contradict this view, and rather suggest some relation with carnivores. As for Hyracoidea, morphologists are undecisive (Novacek 1992) whether to place them with Perissodactyla (Fischer and Tassy 1993; Prothero 1993) or with Proboscidea and Sirenia in the Paenungulata (Simpson 1945; Shoshani 1986). The molecular data, however, unequivocally support their inclusion within the Paenungulata (de Jong, Zweers, and Goodman 1981; Czelusniak et al. 1990; Lavergne et al. 1996; Porter, Goodman, and Stanhope 1996; Stanhope et al. 1996). Morphologically, Tubulidentata is now generally given an unresolved position within the Ungulata (Artiodactyla, Cetacea, Perissodactyla and Paenungulata) (Patterson 1978; Thewissen 1985; Novacek 1992), but a grouping with Insectivora still gets support (Gaudin et al. 1996). Serological data (Shoshani 1986; Sarich 1993), but especially molecular sequences (de Jong, Zweers, and Goodman 1981; Czelusniak et al. 1990; Porter, Goodman, and Stanhope 1996; Stanhope et al. 1996) firmly group aardvark with Paenungulata. Macroscelidea, originally grouped with Insectivora, are now on a morphological basis generally placed as an outgroup of the Glires (Novacek 1992; Gaudin et al. 1996), although a derivation from Condylarthra - the ungulate ancestors - has also been proposed (Hartenberger 1986; Simons, Holroyd, and Brown 1991) (for review, see Butler 1995). Consistent with the latter proposition is the growing molecular evidence that elephant shrews form a monophyletic clade with elephants, sea cows, hyraxes and aardvarks (de Jong, Leunissen, and Wistow 1993; Porter, Goodman, and Stanhope 1996; Stanhope et al. 1996).

To help resolve the disputed affinities of the above four orders, we decided to determine sequences of a suitable protein-coding gene from the nuclear DNA of relevant mammalian species. The aquaporin-2 (AQP2) gene was chosen because it is, in man, a single copy gene - located on chromosome 12 - (Deen et al. 1994a) and demonstrates a promising degree of sequence difference between man (Uchida et al. 1994) and rat (Fushimi et al. 1993). The encoded protein has, moreover, interesting evolutionary and biomedical features. Aquaporins are water channel proteins that facilitate selective water transport across cell membranes in a variety of tissues (Agre, Brown, and Nielsen 1995). AQP2 is exclusively expressed in the kidney collecting duct,

where it is responsible for the vasopressin-dependent concentration of urine (Fushimi et al. 1993; Nielsen et al. 1993; Ma et al. 1994). Mutations in this gene cause nephrogenic diabetes insipidus in man (Deen et al. 1994b; van Lieburg et al. 1994). AQP2 and the other four known mammalian AQPs belong to the ubiquitous major intrinsic protein (MIP) family, which transport small molecules across membranes. These proteins are predicted to consist of six bilayer-spanning α-helices, with the NH_2- and COOH-termini located intracellularly (Gorin et al. 1984; Preston et al. 1994). The protein consists of two repeated halves, supposed to be the result of an intragenic tandem duplication (Wistow, Pisano, and Chepelinsky 1991; Reizer, Reizer, and Saier 1993). The first half is encoded by exon 1 of the AQP2 gene and consists of 121 amino acids.

It is the coding region of this exon 1 that we now have sequenced in 10 mammalian species. Together with the published rat and human sequences, these provide a data set representing 12 eutherian orders. Though relatively short, these sequences contain considerable phylogenetic signal, which has been analyzed, also in combination with sequences of the eye lens protein αA-crystallin from the same taxa.

Material and Methods

Aquaporin-2

The AQP2 sequences reported here are from the following species and orders: nine-banded armadillo (*Dasypus novemcinctus*; Edentata), Indian elephant (*Elephas maximus*; Proboscidea), aardvark (*Orycteropus afer*; Tubulidentata), elephant shrew (*Marcoscelides proboscideus*; Macroscelidea), Cape hyrax (*Procavia capensis*; Hyracoidea), pangolin (*Manis sp*; Pholidota), rabbit (*Oryctolagus cuniculus*; Lagomorpha), horse (*Equus caballus*; Perissodactyla), dog (*Canis familiaris*; Carnivora), and bovine (*Bos taurus*; Artiodactyla). The AQP2 sequences from rat (Fushimi et al. 1993) and human (Uchida et al. 1994) were extracted, as published, from the EMBL database. The primers AQP2FOR (5'-GCA GCA TGT GGG ARC TNM G-3') and AQP2REV (5'-CTY ACI GCR TTI ACN GCN AGR TC-3'), spanning positions –5 to +14 and +343 to +364, respectively, of the first exon of AQP2 were used for PCR (numbers refer to the first nucleotide of the start codon as +1). PCR reactions were performed in 50 μl, con-taining 20 mM Tris-HCl, pH 8.4; 50 mM KCl; 1-2 mM $MgCl_2$; 0.27% Tween 20; 0.27% Nonidet P40; 0.2 mg/ml BSA; 200 μM dNTP (Boehringer Mannheim and Gibco BRL); 0.4 - 5 μM of each primer; 1 U of *Taq* DNA polymerase (provided by Dr. Wiljan Hendriks, University of Nijmegen); and 50–500 ng of genomic DNA as template. Reactions were carried out in a Biometra (Tampa, FL) TRIO-thermoblock using the following program: 7 min at 95 °C, 30 cycles of 2 min at 95 °C, 1 min at 54 - 62 °C and 30 sec at 72 °C, and finally, 5 min at 72 °C. PCR products were ligated directly into a T/A cloning vector (Invitrogen or Promega) according to the manufacturers' instructions, followed by transformation in *E. coli* JM109. Clones containing AQP2 inserts were isolated by means of filter hybridization using an 800-bp fragment of rat AQP2 cDNA as probe. Sequences in both directions were determined from at least three independently obtained PCR products, using the Sequenase version 2.0 sequencing kit (UBS). All sequences have been deposited in the EMBL database under the accession numbers Y10629-Y10638.

αA-Crystallin

Sequences of αA-crystallin used in this study were obtained from the Swissprot database, except the human sequence, which was obtained from the EMBL database, and the elephant shrew (*Macroscelides proboscideus*) sequence, which was as reported in de Jong, Leunissen, and Wistow (1993). Species, orders and accession numbers were: bovine (*Bos taurus*; Artiodactyla), P02470; three-fingered sloth (*Bradypus variegatus*; Edentata), P02487; dog (*Canis familiaris*; Carnivora), P02473; horse (*Equus caballus*; Perissodactyla), P02478; African elephant (*Loxodonta africana*; Proboscidea), P02498; Malayan pangolin (*Manis javanica*; Pholidota), P02498; aardvark (*Orycteropus afer*; Tubulidentata), P02501; Cape hyrax (*Procavia capensis*; Hyracoidea), P02499; rabbit (*Oryctolagus cuniculus*; Lagomorpha), P02493; rat (*Rattus norvegicus*; Rodentia), P02490; human (*Homo sapiens*; Primates) P02489.

Phylogenetic analyses

Sequences were aligned using the PILEUP program from the GCG package (Devereux, Haeberli and Smithies 1984), followed by manual editing to obtain maximum-similarity alignments. DNA data were analyzed with the following methods: (1) the Stationary Markov Model (SMM) (Saccone et al. 1990) to calculate pairwise genetic distances and the NEIGHBOR and DRAWGRAM programs to construct the phylogenetic trees (Felsenstein 1993) and (2) the maximum-likelihood method by using the DNAML program of the PHYLIP package (Felsenstein 1993) assuming transition/transversion ratios of 1.5, 2.0, 2.5, 3.0 and 5.0, and randomizing the order of input four times. Protein data were analyzed as follows: (1) with the maximum-parsimony method by using the PAUP program (Swofford 1993); (2) with the maximum-likelihood method by using the PROTML program (Adachi and Hasegawa 1992); and (3) with the Kimura (1983) method to calculate pairwise genetic distances with the PROTDIST program of the PHYLIP package and the NEIGHBOR and DRAWGRAM programs to construct the phylogenetic trees. The statistical significance of molecular phylogenies was assessed with the bootstrap simulation technique by using the SEQBOOT and CONSENSE programs of the PHYLIP package.

Results and Discussion

Molecular Evolution of Aquaporin-2

Exon 1 of the gene for AQP2 codes for the first 121 amino acids of this protein, which span the N-terminal extension, three transmembrane domains, the interdomain loops A and B, and most of the connecting

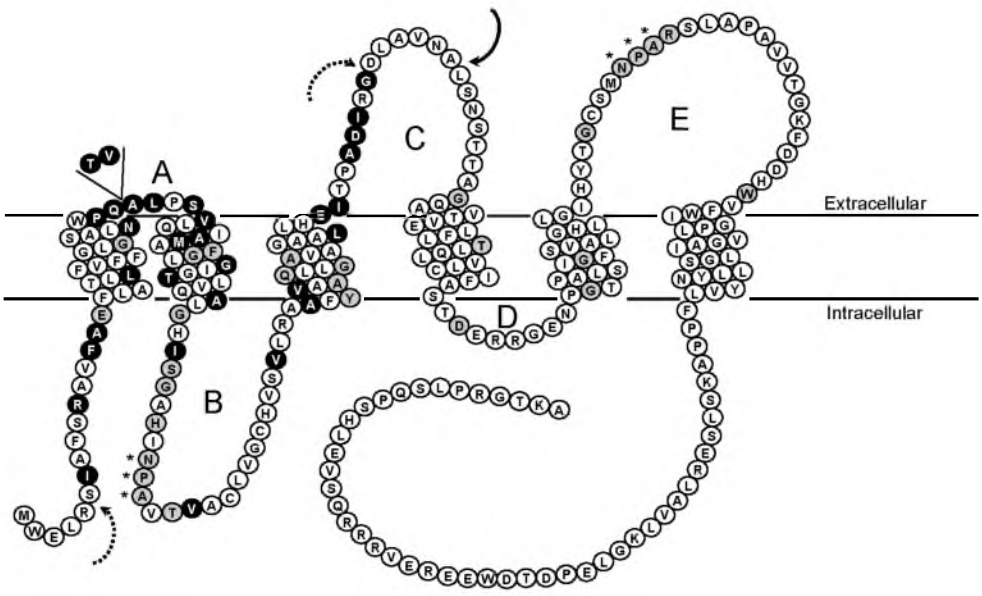


FIG. 1. – Predicted secondary structure of human aquaporin-2. Residues conserved within the mammalian aquaporin family are shown in gray. Residues that vary between the 12 mammalian AQP2 sequences (cf. fig. 3) are black. Black arrow indicates the end of exon 1; dashed arrows demarcate the AQP2 sequence described in this study; asterisks locate the functionally important NPA boxes. The insertion of two residues in loop A of elephant shrew is indicated.

peptide C (fig. 1). Based on the rat and human sequences, degenerated PCR primers were designed to amplify on genomic DNA a 328-bp fragment of the AQP2 exon 1, coding for residues 6-114. Unambiguous PCR products were obtained, cloned and sequenced from 10 mammalian species representing as many orders. The obtained DNA sequences (fig. 2) and the deduced amino acid sequences (fig. 3) were aligned with their rat and human counterparts. The similarity between the different AQP2 proteins varies between 81% and 97%. From figures 1 and 3 it appears that most of the variation at the protein level is concentrated in the extracellular loops A and, as far as determined, C. In loop A, an insertion of two residues occurred in elephant shrew AQP2. Despite this variability, it seems that two prolines are required in loop A, probably to allow a correct bending of this short peptide between two transmembrane helices. As is obvious from figures 1 and 3, none of the conserved residues in the aquaporin family has been replaced in the mammalian AQP2 sequences. This is all in agreement with the structural topology as predicted by the "hourglass" model, where the Asn-Pro-Ala sequence in loop B is assumed to combine in the membrane with a similar "NPA box" in loop E to form the pore responsible for the transport of water (Jung et al. 1994). This model suggests no special function for loops A, C, and D. The hourglass model is derived from studies of AQP1, but also fits for AQP2, as based on functional analyses of AQP2 mutants involved in nephrogenic diabetes insipidus (Deen et al. 1994b; van Lieburg et al. 1994; Mulders et al. 1996). However, recently another structure has been proposed for AQP2 (Bai et al. 1996) which predicts that loops C and D are equally as important as loops B and E in the formation of the aqueous pathway. Although only seven residues of loop C have been compared, it is striking that four out of these seven positions show conspicuous variability. This would argue against an important role of loop C in AQP2, and rather favours the hourglass model.

Phylogenetic Analyses

Aquaporin-2

To study the relationships between eutherian orders, a marsupial would have been the outgroup of choice. However, attempts to amplify the AQP2 gene of an opossum and a kangaroo remained unsuccessful. We therefore resorted to using the armadillo as outgroup, considering that Edentata are generally hypothesized to be the most basal extant eutherian order (McKenna 1975; Novacek 1992; Allard, McNiff, and Miyamoto 1996; but see Gaudin et al. 1996), although molecular data to support this notion are scarce.

Visual inspection of the protein alignment in figure 3 already reveals, apart from the usual homoplasia, the most conspicuous phylogenetic signal. Fourteen phylogenetically informative positions are present. The only occasion where a combination of several unique amino acid replacements occurs in different species is seen in elephant, hyrax, aardvark, and elephant shrew. Three unequivocally apomorphous replacements solely occur in these four species: 15 A→S, 58 A→T and 107 I→L, while also the replacements at position 87 may reflect a shared derived origin. These synapomorphic replace-ments are indeed the result of unique point mutations in the DNA (fig. 2).

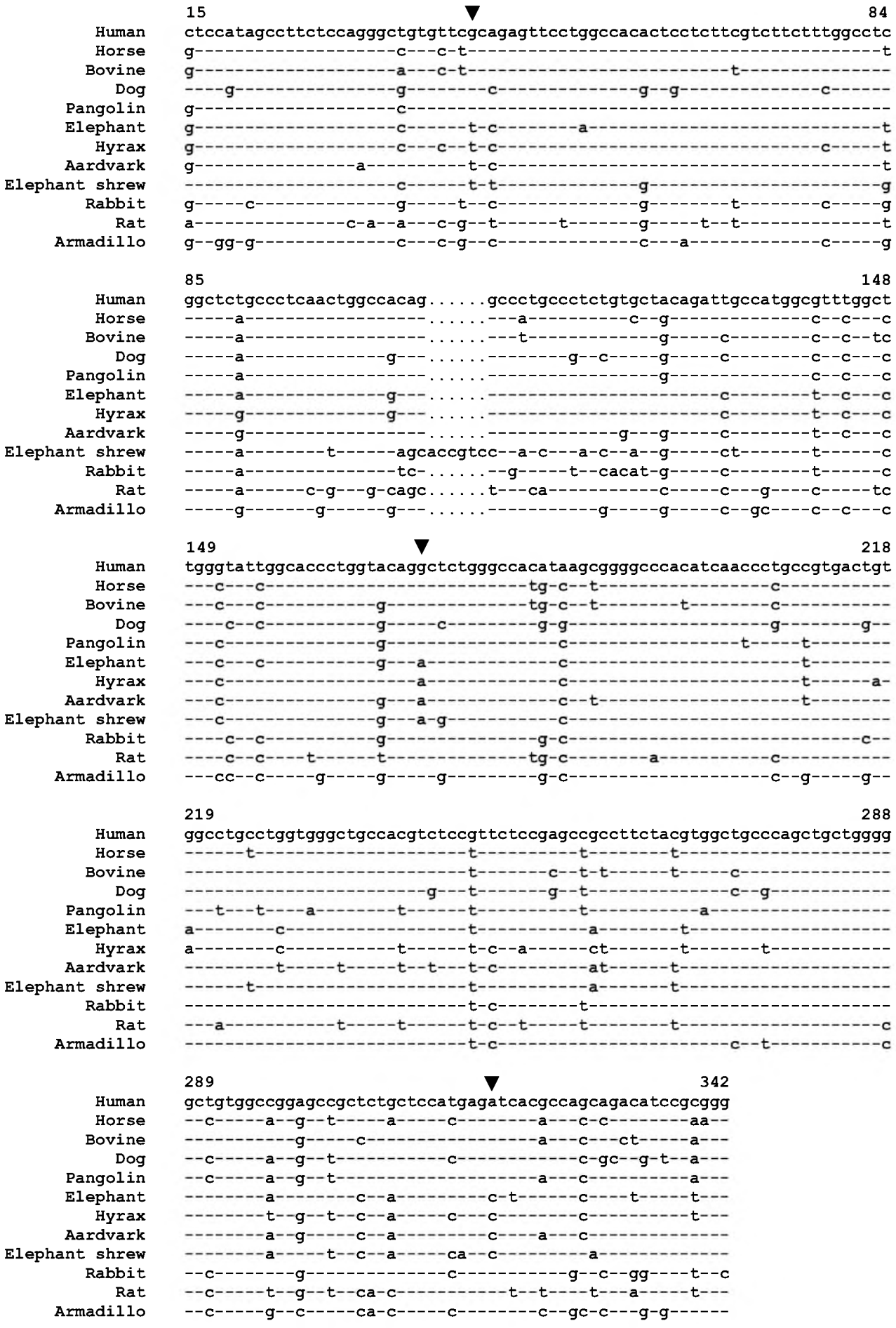


FIG 2.– Alignment of the nucleotide sequences from exon 1 of the AQP2 gene from 12 mammalian species. Position numbers are according to the human sequence (Uchida et al. 1994) with the first nucleotide of the start codon as 1. Hyphens denote nucleotide identities and periods indicate gaps. Arrowheads indicate the positions of three synapomorphous substitutions in aardvark, elephant shrew and paenungulates.

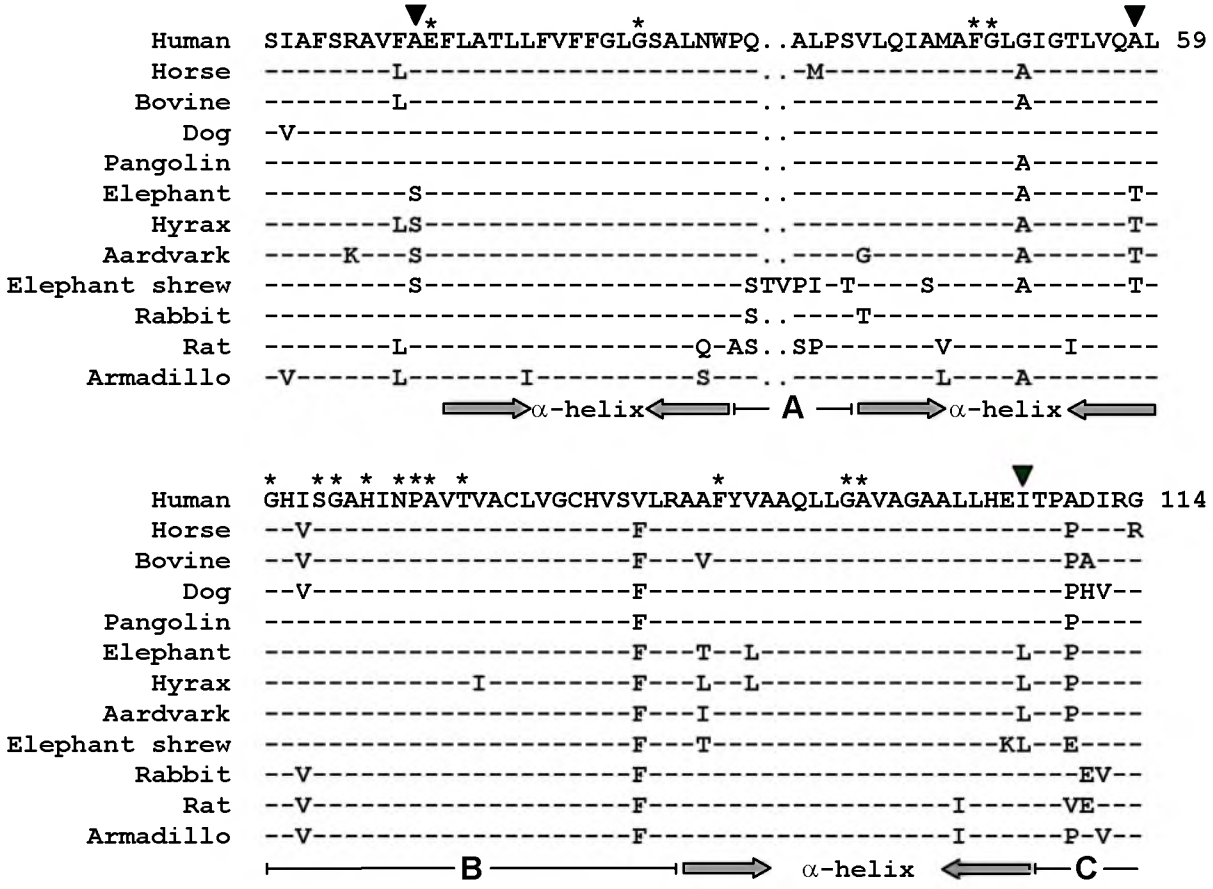


FIG. 3. – Alignment of the deduced amino acid sequences of AQP2 from 12 mammalian species. Position numbers follow the human sequence, with the initiation methionine as 1. Hyphens denote amino acid identities, periods indicate gaps, asterisks show residues conserved throughout the mammalian aquaporin family and arrowheads indicate synapomorphous replacements in aardvark, elephant shrew, and paenungulates. The putative transmembrane helices and connecting loops A-C are indicated (cf. fig. 1).

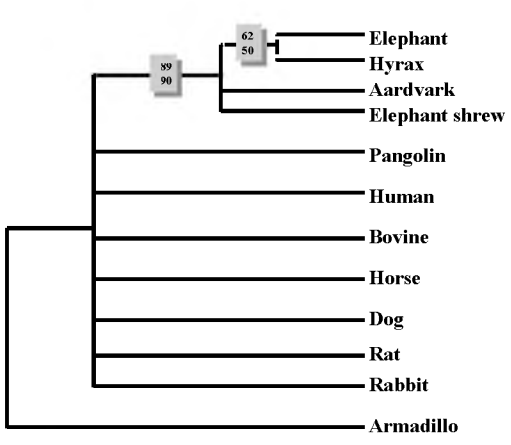


FIG. 4. – Consensus tree, based on the amino acid sequences derived from exon 1 of AQP2 (cf. fig.3). Majority-rule consensus trees were obtained both by the maximum-parsimony method using the PAUP program (Swofford 1993) and by the neighbor-joining method (Saitou and Nei 1987) as applied on a distance matrix calculated according to Kimura (1983). In these trees, all nodes supported by bootstrap values under 50% were collapsed into polytomies, resluting in the same topology for both methods. Bootstrap percentages are based on 1,000 replications and represented for the maximum-parsimony and neighbor-joining methods (top and bottom values, respectively). Armadillo was used as outgroup for both methods.

This observation is fully corroborated by the phylogenetic analyses of the aligned amino acid and nucleotide sequences of AQP2. The amino acid sequences were analyzed by maximum-parsimony and Kimura/neighbor-joining methods (fig. 4). In both cases, the only significant phylogenetic relationship was found to be in the clade consisting of elephant shrew, aardvark, hyrax and elephant, supported by bootstrap values of 89% and 90% for the maximum-parsimony and Kimura/neighbor-joining methods, respectively. The same clade is again the first to be recognized by the maximum-likelihood analysis as carried out by using the program PROTML (Adachi and Hasegawa, 1992):(((elephant, hyrax), aardvark), elephant shrew) (further data not shown). Within this clade, all three methods weakly supported the elephant and hyrax as sister groups. Figure 5 shows the consensus tree similarly obtained by both the SMM/neighbor-joining method and the maximum-likelihood method, on first and second codon positions of AQP2 DNA sequences. The clade including elephant shrew, aardvark, hyrax and elephant is also supported at the DNA level by bootstrap values of 78% and 89% for the SMM/neighbor-joining and maximum-likelihood methods, respectively. No other clade obtained convincing and consistent support from the AQP2 data with all methods.

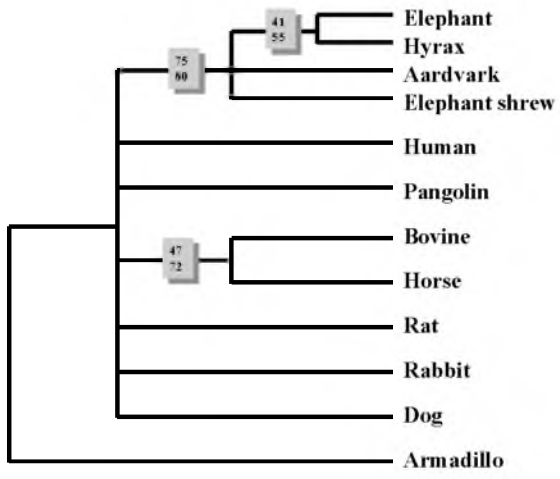


FIG. 5. – Consensus tree, based on first and second codon positions of AQP2 exon 1 DNA sequences (cf. fig. 2). Majority-rule consensus trees were obtained both by the neighbor-joining method (Saitou and Nei 1987) using the distance matrix as calculated with the Stationary Markov Model (Saccone et al. 1990) and by the maximum likelihood method (DNAML with transition/transversion ratio 2.0; Felsenstein, 1981). All nodes supported by bootstrap values under 50% in both these trees were collapsed into polytomies, resulting in the presented topology. Bootstrap percentages are based on 100 replications and represented for the neighbor-joining and maximum-likelihood methods (top and bottom values, respectively). Armadillo was used as outgroup.

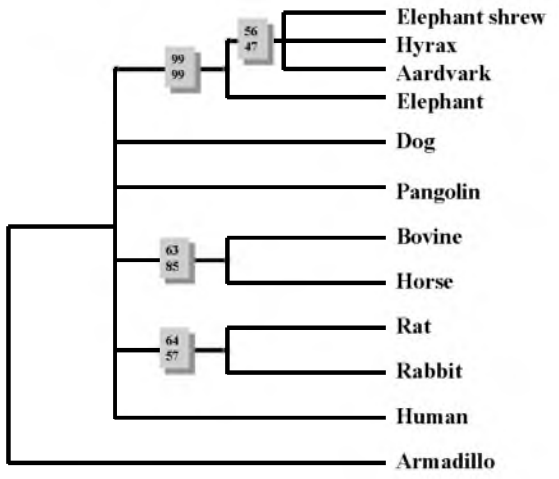


FIG. 6. – Consensus tree, based on the tandemly combined protein seqences of AQP2 exon 1 and αA-crystallin. Majority-rule consensus trees were obtained both by the maximum-parsimony method using the PAUP program (Swofford 1993), and by the neighbor-joining method (Saitou and Nei 1987) applied on a distance matrix calculated by using the method of Kimura (1983). All nodes supported by bootstrap values under 50% in both trees were collapsed into polytomies. Bootstrap percentages as based on 1,000 replications, are represented for the maximum-parsimony and neighbor-joining methods (top and bottom, respectively). Edentates was used as outgroup.

Combined Aquaporin-2 and αA-Crystallin

The eye lens protein αA-crystallin is the only other protein for which sequences are known from the same 12 orders (including Pholidota) as described for AQP2. In order to see if additional phylogenetic information could be retrieved, the AQP2 and αA-crystallin amino acid sequences representing these 12 orders were tandemly combined, aligned, and subsequently subjected to the same phylogenetic analyses as the AQP2 amino acid sequence data. The combined amino acid sequences together have a length of 282 amino acids, except for the combined elephant shrew sequence (length 248 residues) and the edentate sequence (combined armadillo AQP2 and three-fingered sloth αA-crystallin, length 279 residues). Phylogenetic analyses of the tandemly aligned data set even more strongly supported a close phylogenetic relationship between elephant shrew, aardvark, hyrax, and elephant, with bootstrap values of 99% for both the maximum- parsimony method and for the Kimura/neighbor-joining method (fig. 6). The parsimony analysis required 102 steps and gave only one tree. Both methods also weakly supported some other superordinal clades, notably those grouping bovine and horse, and rat and rabbit. Also, the maximum-likelihood (PROTML) analysis on this combined data set supported the elephant/hyrax/elephant shrew/aardvark clade in all of the 32 top ranking trees, while the horse/cow and rat/rabbit clades received support in 28 and 25 of the 32 trees, respectively (data not shown).

Paenungulate Phylogeny

Molecular sequence evidence showing that not only aardvark but - surprisingly - also elephant shrews should be grouped in one clade with the paenungulates was first obtained from the eye lens protein αA-crystallin (de Jong, Zweers, and Goodman 1981; de Jong, Leunissen, and Wistow 1993). This evidence is now fully supported by the AQP2 sequences and even more strongly supported by the recent studies of exon 28 of the von Willebrand factor (vWF) (Porter, Goodman, and Stanhope 1996) and of exon 1 of interphotoreceptor retinoid binding protein (IRBP) (Stanhope et al. 1996). Within this clade, the AQP2, vWF and IRBP data all indicate that hyrax is more closely related to Tethytheria (elephants and sirenians; McKenna 1975) than are aardvark and elephant shrew. αA-Crystallin, and as a consequence the combined AQP2/αA-crystallin protein analyses, place elephant as the first branch off the clade, probably as the result of a single back mutation (de Jong, Leunissen, and Wistow 1993). The branching order of elephant shrew and aardvark is less clear, although the analyses of especially vWF, but also AQP2 and - depending on the method - IRBP, suggest that aardvark is closer to elephant and hyrax than is elephant shrew.

The various paleontological and morphological evidence to support the widely diverging opinions about the relationships of hyrax, aardvark, and elephant shrew has been amply discussed by Porter, Goodman, and Stanhope (1996) and Stanhope et al. (1996). Molecular sequences had until recently contributed little to solving these phylogenetic problems. Four data sets are now available that include these orders as well as a sufficient number of other eutherian orders. The fact that these data sets

consistently and independently provide strong support for a close relationship of aardvark and elephant shrew with Paenungulata leaves little doubt that this clade represents one of the most conspicuous lineages in eutherian radiation. Evidence from a single nuclear gene can easily be dismissed by assuming hidden paralogy, or invoking a molecular evolutionary mechanism like gene conversion or homoplasia due to functional constraints. Such explanations can not apply, however, to the genes coding for proteins with such diverse structures and functions as αA-crystallin, vWF, IRBP and AQP2. It is puzzling why this pronounced phylogenetic signal at the molecular level is apparently not paralleled by similarly convincing paleontological and morphological evidence.

Pholidote Relationships

Although no other phylogenetic relationships gained consistent support from the AQP2 data itself or in combination with αA-crystallin, some trends are noteworthy. AQP2 and αA-crystallin are the only available sequences yet for pangolins. αA-Crystallin alone placed pangolin within an unresolved ensemble of artiodactyls, cetaceans, carnivores and perissodactyls (de Jong and Goodman 1982; de Jong, Leunissen, and Wistow 1993). It actually was depicted as the sistergroup of bears because of a possibly homoplastic replacement. However, the pangolin was clearly far removed from edentates, which branched off at the root of the eutherians. Similarly, the AQP2 data do not support a sister group relationship of pholidotes and edentates, in agreement with the considerable and growing morphological evidence that contradicts such a grouping (McKenna 1975; Rose and Emry 1993). However, the combined AQP2 and αA-crystallin evidence (fig. 6) does not significantly support clustering of pangolin with any other order either.

Other Superordinal Clades

Artiodactyls and perissodactyls are associated in the SMM/neighbor-joining and maximum-likelihood analyses of the AQP2 DNA sequences (fig. 5), as well as in maximum-parsimony and Kimura/neighbor-joining analyses of the combined AQP2/αA-crystallin protein sequences (fig. 6). Perissodactyls are also placed closest to the artiodactyl/cetacean clade in the IRBP analyses (Stanhope et al. 1996) and in earlier analyses of five to seven combined protein sequences (Czelusniak et al. 1990). Also, the maximum-parsimony analysis of the combined mitochondrially encoded proteins from species representing up to seven eutherian orders supports this grouping, but neighbor-joining analyses of these proteins and of the RNA-coding mitochondrial genes rather place perissodactyls and carnivores as sister groups (D'Erchia et al. 1996; Xu, Janke, and Arnason 1996). Neither grouping is in agreement with the prevailing morphological opinion, which places Perissodactyla closer to the Paenungulata (Novacek 1992; Prothero 1993).

Finally, it is remarkable that the cohort Glires gets some support in the analyses of the combined AQP2 and αA-crystallin sequences (fig. 6). Also the IRBP data set gave some evidence for such a grouping of rodents and lagomorphs (Stanhope et al. 1996), which is otherwise strongly refuted by most molecular analyses (Graur, Duret, and Gouy 1996; D'Erchia et al. 1996).

This study shows that even a relatively short sequence - 328 bp from the first exon of the AQP2 gene - may provide significant phylogenetic signal. It is a further indication that there is no reason anymore for Simpson's (1945) pessimistic view that the eutherian radiation would be irresolvable. It is clear, however, that many more nuclear genes, as well as complete mitochondrial genomes, need to be analyzed in a sufficiently broad sampling of species to convincingly resolve the higher level mammalian phylogeny.

Acknowledgments

For tissue samples, we thank Drs. P.R. Klaster, Royal Tropical Institute, Amsterdam, The Netherlands (armadillo), J. Wensing, Burgers Zoo, Arnhem, The Netherlands (aardvark), G. Olbricht, Wuppertal Zoo, Germany (elephant shrew), E. Harley, South Africa (hyrax) and G. Bernardi, Paris, France (pangolin) and the Noorder Zoo, Emmen, The Netherlands (elephant). We thank Rogier van Otterlo for technical assistance. This study was supported by grants from the European Commission (HCM CHRXCT930254) to W.W. de J. and C.S., and from the Dutch Kidney Foundation to P.M.T.D. (93.1299).

LITERATURE CITED

Adachi, J., and M. Hasegawa. 1992. MOLPHY: programs for molecular phylogenetics. I. PROTML: maximum likelihood inference of protein phylogeny. Computer Science Monographs, no. **27**. Institute of Statistical Mathematics, Tokyo.

Agre, P., D. Brown, and S. Nielsen. 1995. Aquaporin water channels: unanswered questions and unresolved controversies. Curr. Opin. Cell. Biol. **7**:472-483.

Allard, M.W., B.E. McNiff, and M.M. Miyamoto. 1996. Support for interordinal eutherian relationships with an emphasis on primates and their archontan relatives. Mol. Phylogenet. Evol. **5**:78-88.

Archibald, J.D. 1996. Fossil evidence for a late cretaceous origin of "hoofed" mammals. Science **272**:1150-1153.

Bai, L., K. Fushimi, S. Sasaki, and F. Marumo. 1996. Structure of aquaporin-2 vasopressin water channel. J. Biol. Chem. **271**:5171-5176.

Benton, M.J. ed. 1988. The Phylogeny and Classification of the Tetrapods Vol 2. Clarendon Press, Oxford.

Butler, P.M. 1995. Fossil Macroscelidea. Mamma. Rev. **25**: 3-14.

Czelusniak, J., M. Goodman, B.F. Koop, D.A. Tagle, J. Shoshani, G. Braunitzer, T.K. Kleinschmidt, W.W. de Jong, and G. Matsuda. 1990. Perspectives from amino acid and nucleotide sequences on cladistic relationships among higher taxa of Eutheria. Pp. 545-572 *in* H.H. Genoways, ed. Current Mammalogy Vol 2. Plenum Press, New York.

Deen, P.T.M., D. Olde Weghuis, R.J. Sinke, A. Geurts van Kessel, B. Wieringa, and C.H. van Os. 1994a. Assignment of the human gene for the water channel of renal collecting duct aquaporin 2 (AQP2) to chromosome 12, region q12-q13. Cytogenet. Cell. Genet. **66**:260-262.

Deen, P.T.M., M.A.J. Verdijk, N.V.A.M. Knoers, B. Wieringa, L.A.H. Monnens, C.H. van Os, and B.A. van Oost. 1994b. Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. Science **264**:92-95.

de Jong, W.W., A. Zweers, and M. Goodman. 1981. Relationship of aardvark to elephants, hyraxes and sea cows from α-crystallin sequences. Nature **292**:538-540.

de Jong, W.W., and M. Goodman. 1982. Mammalian phylogeny studied by sequence analysis of the eye lens protein α-crystallin. Z. f. Säugetierk **5**:257-276.

de Jong, W.W., J.A.M. Leunissen, and G.J. Wistow. 1993. Eye lens crystallins and the phylogeny of placental orders: evidence for a macroscelid-paenungulate clade? Pp. 5-12 *in*: F.S Szalay, M.J. Novacek, and M.C. McKenna, eds. Mammal Phylogeny: Placentals. Springer Verlag, New York.

D'Erchia, A.M., C. Gissi, G. Pesole, C. Saccone, and U. Arnason. 1996. The guinea pig is not a rodent. Nature **381**:597-600.

Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic. Acids Res. **12**:387-395.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368-376.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Fischer, M.S., and P. Tassy. 1993. The interrelation between Proboscidea, Sirenia, Hyracoidea, and Mesaxonia: the morphological evidence. Pp. 217-234 *in* F.S. Szalay, M.J. Novacek, and M.C. McKenna, eds. Mammal Phylogeny: Placentals. Springer Verlag, New York.

Fushimi, K., S. Uchida, Y. Hara, Y. Hirata, F. Marumo, and S. Sasaki. 1993. Cloning and expression of apical membrane water channel of rat kidney collecting tubule. Nature **361**:549-552.

Gaudin, T.J., J.R. Wible, J.A. Hopson, and W.D. Turnbull. 1996. Reexamination of the morphological evidence for the cohort Epitheria (Mammalia, Eutheria). J.Mammal. Evol. **3**:31-79.

Gorin, M.B., S.B. Yancey, J. Cline, J.P. Revel, and J. Horwitz. 1984. The major intrinsic protein (MIP) of lens fiber membrane. Cell **39**:49-59.

Graur, D. 1993. Towards a molecular resolution of the ordinal phylogeny of the eutherian mammals. FEBS **325**:152-159.

Graur, D., L. Duret, and M. Gouy. 1996. Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). Nature **379**:333-335.

Hartenberger, J.L. 1986. Hypothese paleontologique sur l'origine des Macroscelidea (Mammalia). C. R. Acad. Sci. Ser. II **302**:247-249.

Hedges, S.B., P.H. Parker, C.G. Sibley, and S. Kumar. 1996. Continental breakup and the ordinal diversification of birds and mammals. Nature 381:226-229.

Honeycutt, R.L., and R.M. Adkins. 1993. Higher level systematics of eutherian mammals: an assessment of molecular characters and phylogenetic hypotheses. Annu. Rev. Ecol. Syst. **24**:279-305.

Jung, J.S., G.M. Preston, B.J. Smith, W.B. Guggino, and P. Agre. 1994. Molecular structure of the water channel through aquaporin CHIP. J. Biol. Chem. **269**:14648-14654.

Kimura, M., (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Lavergne, A., E. Douzery, T. Stichler, F.M. Catzeflis, and M.S. Springer. 1996. Interordinal mammalian relationship: evidence for paenungulate monophyly is provided by complete mitochondrial 12S rRNA sequences. Mol. Phylogenet. Evol. **6**:245-258.

Ma, T., H. Hasegawa, W.R. Skach, A. Frigeri, and A.S. Verkman. 1994. Expression, functional analysis, and in situ hybridization of a cloned rat kidney collecting duct water channel. Am. J. Physiol. **266**:C189-197.

McKenna, M.C. 1975. Toward a phylogenetic classification of the Mammalia. Pp. 21-46 *in*: W.P. Luckett, and F.S. Szalay, eds. Phylogeny of the Primates: An Interdisciplinary Approach. Plenum Press, New York.

Mulders, S.M., N.V.A.M. Knoers, A.F. van Lieburg, L.A.H. Monnens, E. Leumann, K. Schärer, E. Schober, J.P.L. Rijss, C.H. van Os, and P.M.T. Deen. 1996. New mutations in the AQP2 gene in nephrogenic diabetes insipidus resulting in functional but misrouted water channels. J. Clin. Invest. (in press).

Nielsen, S., S.R. DiGiovanni, E.I. Christensen, M.A. Knepper, and H.W. Harris. 1993. Cellular and subcellular immunolocalization of vasopressin regulated water channel in rat kidney. Proc. Natl. Acad. Sci. USA **90**:11663-11667.

Novacek, M.J. 1992. Mammalian phylogeny: shaking the tree. Nature **356**:121-125.

Patterson, B. 1978. Pholidota and Tubulidentata. Pp. 268-278 *in* V.J. Maglio, and H.B.S Cooke. eds. Evolution of African Mammals. Harvard University Press, Cambridge, Massachusetts.

Porter, C.A., M. Goodman, and M.J. Stanhope. 1996. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gene. Mol. Phylogenet. Evol. **5**:89-101.

Preston, G.M., J.S. Jung, W.B. Guggino, and P. Agre. 1994. Membrane topology of aquaporin CHIP. J. Biol. Chem. **269**:1668-1673.

Prothero, D.R. 1993. Ungulate phylogeny: molecular vs. morphological evidence. Pp. 173-181 *in* F.S Szalay, M.J. Novacek, and M.C. McKenna, eds. Mammal Phylogeny: Placentals. Springer Verlag, New York.

Reizer, J., A. Reizer, and M.H. Saier. 1993. The MIP family of integral membrane proteins: sequence comparisons, evolutionary relationships, reconstructed pathway of evolution, and proposed functional differentiation of the two repeated halves of the proteins. Crit. Rev. Biochem. Mol. Biol. **28**:235-257.

Rose, K.D., and R.J. Emry. 1993. Relationship of Xenarthra, Pholidota, and fossil "Edentates": The morphological evidence. Pp. 81-101 *in*: F.S Szalay, M.J. Novacek, and M.C. McKenna, eds. Mammal Phylogeny: Placentals. Springer Verlag, New York.

Saccone, C., C. Lanave, G. Pesole, and G. Preparata. 1990. Influence of base composition on quantitative estimates of gene evolution. Methods Enzymol. **183**:570-583.

Saitou, N., and M. Nei. 1987. The neighbour-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406-425.

Sarich, V.M. 1993. Mammalian systematics: twenty-five years among their albumins and transferrins. Pp. 103-114 *in* F.S. Szalay, M.J. Novacek, and M.C. Mckenna, eds. Mammal Phylogeny: Placentals. Springer Verlag, New York.

Shoshani, J. 1986. Mammalian phylogeny: comparison of morphological and molecular results. Mol. Biol. Evol. **3**:222-242.

Simons, E.L., P.A. Holroyd, and T.M. Brown. 1991. Early tertiary elephant shrews from Egypt and the origin of the Macroscelidea. Proc. Natl. Acad. Sci. USA **88**:9734-9737.

Simpson, G.G. 1945. The principles of classification and a classification of mammals. Bull. Am. Mus. Nat. Hist. **85**:1-350.

Stanhope, M.J., M.R. Smith, V.G. Waddell, C.A. Porter, M.S. Shivij, amd M. Goodman. 1996. Mammalian evolution and the IRBP gene: convincing evidence for several superordinal clades. J. Mol. Evol. **43**:83-96.

Swofford, D.L. 1993. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1, Computer program distrubuted by the Illinois Natural History Survey, Champaign.

Szalay, F.S., M.J. Novacek, and M.C. McKenna eds. 1993. Mammal Phylogeny: Placentals. Springer Verlag, New York.

Thewissen, J.G.M. 1985. Cephalic evidence for the affinities of Tubulidentata. Mammalia **49**:257-284.

Uchida, S., S. Sasaki, K. Fushimi, and F. Marumo. 1994. Isolation of human aquaporin-CD gene. J. Biol. Chem. **269**:23451-23455.

van Lieburg, A.F., M.A.J. Verdijk, V.V.A.M. Knoers, A.J. van Essen, W. Proesmans, R.Mallmann, L.A.H. Monnens, C.H. van Oost, and P.M.T. Deen. 1994. Patients with autosomal nephrogenic diabetes insipidus homozygous for mutations in the aquaporin 2 water-channel gene. Am. J. Hum. Genet. **55**:648-652.

Wistow, G.J., M.M. Pisano, and A.B. Chepelinsky. 1991. Tandem sequence repeats in transmembrane channel proteins. Trends Biochem. Sci. **16**: 170-171.

Xu, X., A. Janke, and U. Arnason. 1996. The complete mitochondrial DNA sequence of the greater Indian rhinoceros, *Rhinoceros unicornis*, and the phylogenetic relationship among Carnivora, Perissodactyla, and Artiodactyla (+ Cetacea). Mol. Biol. Evol. **13**:1167-1173

Chapter Three

Endemic African mammals shake the phylogenetic tree

Mark S. Springer, Gregory C. Cleven, Ole Madsen, Wilfried W. de Jong, Victor G. Waddell, Heather M. Amrine and Michael J. Stanhope. 1997. Endemic African mammals shake the phylogenetic tree. *Nature* **388**: 61-64

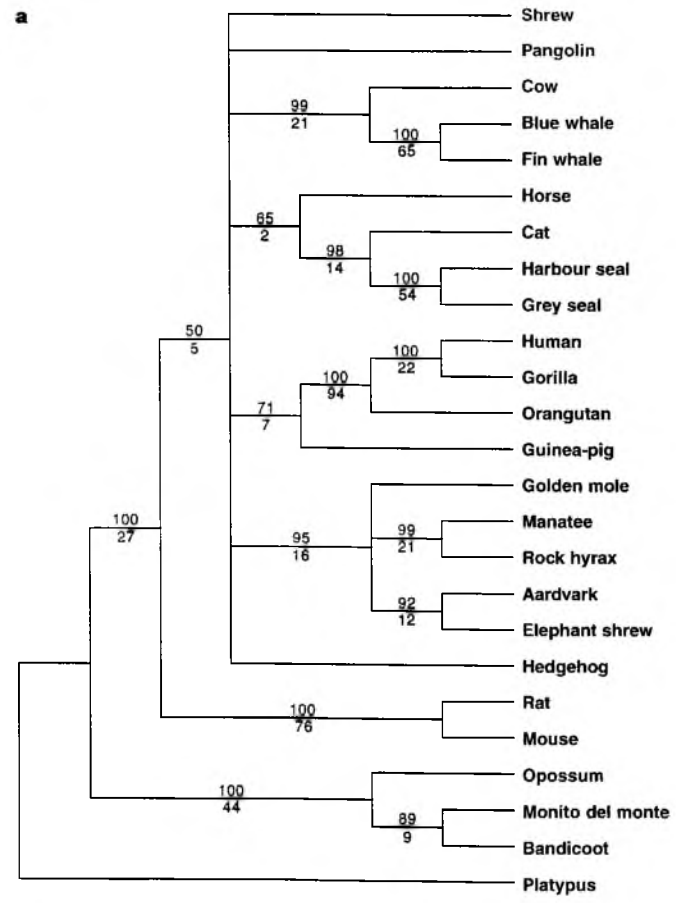
The order Insectivora, including living taxa (lipotyphlans) and archaic fossil forms, is central to the question of higher-level relationships among placental mammals[1]. Beginning with Huxley[2], it has been argued that insectivores retain many primitive features and are closer to the ancestral stock of mammals than are other living groups[3]. Nevertheless, cladistic analysis suggests that living insectivores, at least, are united by derived anatomical features[4]. Here we analyse DNA sequences from three mitochondrial genes and two nuclear genes to examine relationships of insectivores to other mammals. The representative insectivores are not monophyletic in any of our analyses. Rather, golden moles are included in a clade that contains hyraxes, manatees, elephants, elephant shrews and aardvarks. Members of this group are of presumed African origin[5,6]. This implies that there was an extensive African radiation from a single common ancestor that gave rise to ecologically divergent adaptive types. 12S ribosomal RNA transversions suggest that the base of this radiation occurred during Africa's window of isolation in the Cretaceous period before land connections were developed with Europe in the early Cenozoic era.

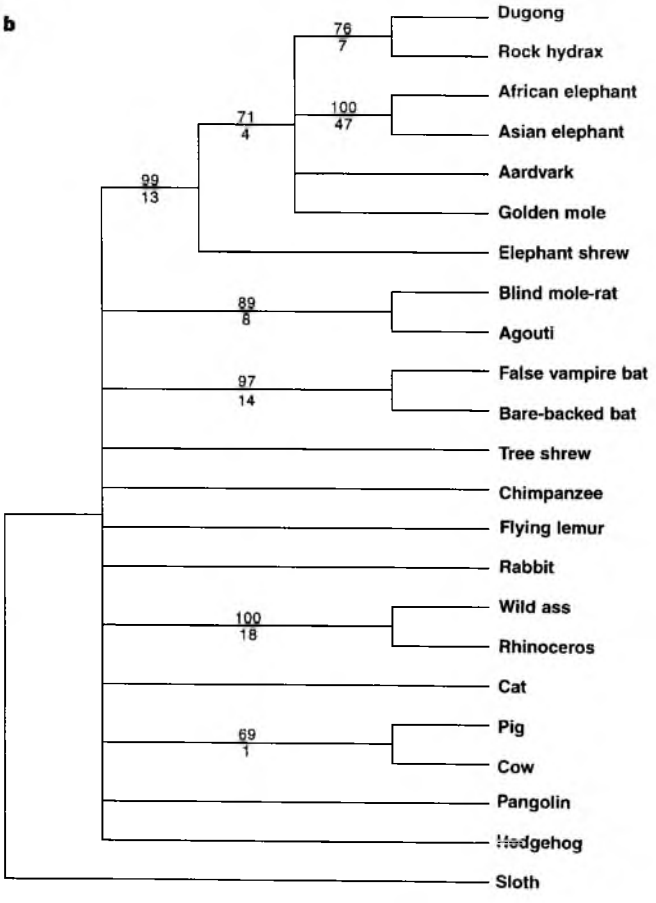
Relationships among orders of placental mammals have proved difficult to resolve[1]. To extend the available mitochondrial (mt) sequences, a 2.6-kilobase (kb) segment containing the 12S rRNA, valine transfer RNA, and 16S rRNA genes was sequenced for nine taxa to generate a data set that is representative of 12 of the 18 placental orders and all three insectivore suborders[4]. Phylogenetic analyses provide strong support for well-established mammalian clades such as carnivores, hominoids, and Cetacea plus Artiodactyla (Fig. 1a). In agreement with other molecular studies[7–10] that included an assortment of taxa, most interordinal associations are not resolved at bootstrap values >75%. However, the mtDNA data do provide strong support for the association of the two paenungulates (hyrax, manatee) together, and of these with elephant shrews, aardvarks and golden moles (Fig. 1a and Table 1). The association of hyraxes with proboscideans and sirenians was suggested by Cope[11]. A competing hypothesis is an association of hyraxes with perissodactyls[12]. Our results agree with earlier protein[13,14] and DNA studies[7–10] supporting Cope's paenungulate hypothesis. In addition to bootstrap support, T-PTP[15] and Kishino–Hasegawa[16] tests also support paenungulate monophyly (Table 2). Anatomical data provide some evidence that aardvarks and/or elephant shrews may be related to paenungulates[17,18] but suggest other hypotheses as well: for example, six osteological features are putative synapomorphies uniting elephant shrews with lagomorphs and rodents[19]. All the available sequence data, including amino-acid sequences[13,14], DNA sequences for three nuclear genes[8–10], and the present mitochondrial genes, support an association of aardvarks and elephant shrews with paenungulates. What is most unexpected is that golden moles, a family of insectivores, are also part of this clade. 12S rRNA sequences earlier suggested an association of golden moles with paenungulates, but did not provide convincing bootstrap support for this hypothesis[7]. Our expanded data set demonstrates that insectivores are not monophyletic (Table 2)

Table 1 Bootstrap support for select clades based on different methods

	Clade	
	Paenungulata	Paenungulata + aardvark + elephant shrew +golden mole
Mitochondrial DNA		
Parsimony	99	95
Transversion parsimony	64	90
Minimum evolution		
Tamura-Nei I	100	92
Tamura-Nei II	100	78
Logdet	99	90
Maximum likelihood	100	100
vWF		
Parsimony		
All positions	49	99
1st and 2nd positions	24	65
3rd positions	51	93
Transversion parsimony	30	95
Minimum evolution		
Tamura-Nei I	37	99
Tamura-Nei II	30	99
Logdet	43	97
Maximum likelihood	78	100
A2AB		
Parsimony		
All sites	71	88
1st and 2nd positions	49	81
3rd positions	31	67
Transversion parsimony	71	54
Minimum evolution		
Tamura-Nei I	83	84
Tamura-Nei II	28	25
Logdet	79	78
Maximum likelihood	81	89

Only two of the three paenungulate orders were represented among the mitochondrial and A2AB sequences. Tamura-Nei[27] I and II distances were calculated by using an equal-rates assumption and a gamma-distribution of rates, respectively.





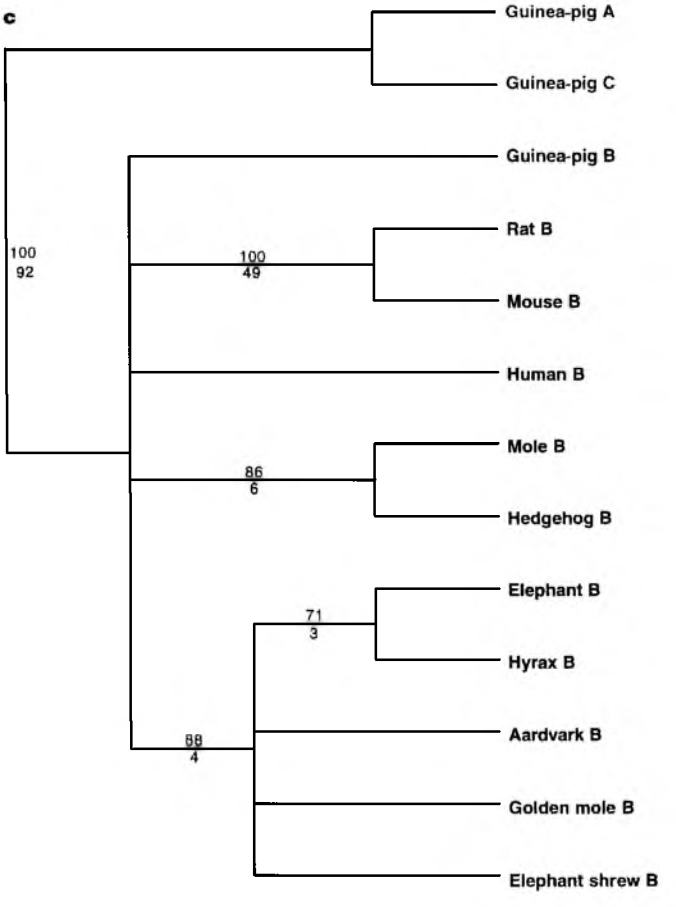


Figure 1 Majority-rule parsimony bootstrap trees based on mitochondrial (**a**), vWF (**b**), and A2AB (**c**) sequences. Bootstrap values and decay indices, respectively, are above and below branches. The minimum-length tree (3,836 steps; consistency index, 0.430) for the mitochondrial data is 4, 7, 43 and 66 steps shorter than trees that constrain shrew + hedgehog, rodent monophyly, insectivore monophyly, and hyrax + horse, respectively. Insectivore monophyly and a hyrax-perissodactyl association require 29 and 64 additional steps, respectively, in comparison to five minimum-length vWF trees (2,401 steps; consistency index, 0.511). Insectivore monophyly requires 26 additional steps in comparison to the shortest A2AB tree (1,469 steps; consistency index, 0.730).

Table 2 Significance levels of T-PTP and Kishino–Hasegawa tests

Constraint	Mitochondrial DNA			vWF			A2AB		
	T-PTP	KH-P	KH-L	T-PTP	KH-P	KH-L	T-PTP	KH-P	KH-L
Perissodactyls + hyracoids	0.01	0.0011–0.0022	<0.0001	0.00	<0.0001	<0.0001	MT	MT	MT
Insectivore monophyly	0.05	<0.0001	0.0001	0.01	0.0311	0.0477	0.00	0.0002–0.0067	0.0001

In each case, trees with constraints were compared against either minimum length (T-PTP, KH-P) or highest likelihood (KH-L) trees. T-PTP tests were based on 100 permutations. KH-P, Kishino–Hasegawa test with parsimony; KH-L, Kishino–Hasegawa test with maximum likelihood; MT, missing taxa.

Table 3 Divergence times (Myr) based on 12S rRNA transversions

Divergence event	*N*	Mean	Standard deviation	Standard error
Among Paenungulates	3	54.8	4.2	2.4
Paenungulates to golden mole	3	67.1	8.7	5.0
Paenungulates to aardvark	3	74.0	12.0	6.9
Paenungulates to elephant shrew	3	79.9	9.9	5.7
African clade to other 13 orders	78	91.1	15.5	1.6

and that golden moles, elephant shrews, aardvarks and paenungulates are part of the same clade.

To corroborate these findings, we obtained sequences from exon 28 of the von Willebrand factor (vWF) gene for golden mole, hedgehog and pangolin. Adding these to the existing vWF data set[9], we found high bootstrap support for the inclusion of golden moles with paenungulates, elephant shrews and aardvarks (Fig. 1b and Table 1). Sequences from the α-2B adrenergic receptor gene (*A2AB*) also support the association of golden moles with paenungulates, elephant shrews and aardvarks (Fig. 1c and Table 1). Parsimony and maximum-likelihood trees supporting the paenungulate–golden mole–aardvark–elephant shrew clade are significantly better than the best trees that constrain insectivore monophyly (Table 2).

This expanded clade, which includes five placental orders plus golden moles, has not been previously hypothesized on the basis of morphological or molecular data. Elephants, sirenians, hyraxes, golden moles, aardvarks and elephant shrews show a variety of ecological and morphological specializations and it is not surprising that morphology has not elucidated their common ancestry, now evident from DNA sequences. It is notable that all six of these groups are of probably African origin or, in the case of the aquatic sirenians, from along the margins of the former Tethys Sea[5,6,13]. In two cases (golden moles and elephant shrews), geographic distribution has been restricted to Africa for the complete temporal range of these taxa[5]. Thus geographic evidence adds to the molecular data in support of this 'African origin' clade. The radiation of the African clade parallels endemic radiations of other vertebrate taxa on Southern Hemisphere continents during the breakup of Gondwanaland; for example, marsupials and passerine birds in Australia[20], and marsupials, edentates and notoungulates in South America[5].

Paenungulate orders diverged from each other 51 to 59 million years (Myr) ago, as deduced from 12S rRNA transversions (Table 3). Deeper in the African clade, average divergence times between paenungulates and other lineages range from 67 to 80 Myr. The mean divergence time between taxa in the African clade and the other 13 orders of placental mammals is ~91 Myr. These divergence times support the hypothesis that many eutherian orders arose before the extinction of dinosaurs at the end of the Cretaceous[21] and imply that conventional views on the origins of the African mammal fauna[5] are incorrect. Africa's window of isolation extended from the Late Cretaceous, when Africa became separated from South America, to the early Cenozoic, when tenuous connections developed between northern Africa and Europe. The window of isolation extended from at least 80 Myr (ref. 20), if not earlier, until the early Cenozoic. The traditional view is that condylarths, prosimian primates and creodont carnivores reached Africa from the north after the docking of Africa with Europe[5]. From the condylarth stock, groups such as proboscideans and sirenians ostensibly originated in Africa. Other elements of the African mammal fauna, including perissodactyls, artiodactyls, insectivores and living carnivore families, presumably arrived in the Neogene with the establishment of the Arabian Peninsula. Evidence for an extensive African clade, including taxa with divergence times as old as 80 Myr, is inconsistent with this view. The ancestor of the African clade probably resided in Africa before the window of isolation and did not arrive from the north in the early Cenozoic. The role of geographic isolation and continental break-up in the early diversification of placental mammals is potentially more important than previously recognized.

Methods

Amplification and sequencing. 12S rRNA and tRNA genes were amplified and sequenced as described[22]. 16S rRNA genes were amplified using primers for valine-tRNA (for example, 5′-tacaccyaraagatttca-3′) and leucine-tRNA (for example, 5′-agaggrtttgaacctctg-3′) and sequenced. Accession numbers for the new mitochondrial sequences (*Echymipera kalubu* (bandicoot); *Dromiciops gliroides* (monito del monte); *Sorex palustris* (shrew); *Manis* sp. (pangolin); *Amblysomus hottentotus* (golden mole); *Procavia capensis* (hyrax); *Trichechus manatus* (manatee); *Orcyteropus afer* (aardvark); *Elephantulus rufescens* (elephant shrew)) are U97335–U97343. 12S rRNA sequences for several of these taxa have been deposited in GenBank (M95108 (golden mole), U61073 (monito del monte), U61079 (pangolin), U61083 (manatee), U61084 (hyrax)). Accession numbers for additional mitochondrial sequences are as follows: cow (J01394); blue whale (X72204); fin whale (X61145); horse (X79547); cat (U20753) harbour seal (X63726); grey seal (X72004); human (J01415); gorilla (D38114); orang-utan (D38115); guinea-pig (L35585); hedgehog (X88898); rat (X14848); mouse (J01420); opossum (Z29573); platypus (U33498; X83427). Exon 28 of the vWF gene was amplified and sequenced as described[9]. Accession numbers for *Manis* sp., *Erinaceus europaeus* (hedgehog), and *Amblysomus hottentotus* vWF sequences are U97534–U97536. Additional vWF sequences are from ref. 9. Part of the single-copy, intronless *A2AB* gene was amplified using the primers A2ABFOR (5′-ascctactcngtgcaggcnacng-3′) and A2ABREV (5′-ctgttgcagtagccdatccaraaraaraaytg-3′). PCR products were cloned into a T/A cloning vector (Promega) and both strands were sequenced for at least two clones using the Thermo Sequenase fluorescent-labelled primer cycle sequencing kit (Amersham). Accession numbers for the new *A2AB* sequences (*Elephas maximus* (elephant); *Orycteropus afer* (aardvark); *Macroscelides proboscideus* (elephant shrew); *Amblysomus hottentotus* (golden mole); *Procavia capensis* (hyrax); *Erinaceus europaeus* (hedgehog); *Talpa europaea* (mole)) are Y12520–Y12526. Additional α-2 adrenergic sequences are M34041 (human); M32061 (rat); (L00974) (mouse), and U25722–U25724 (guinea-pig).

Sequence alignment and phylogenetic analysis. Sequences were aligned using CLUSTAL W (ref. 23). rRNA alignments were modified in view of secondary structure[24,25]. Ambiguous regions were omitted from subsequent analyses[26]; this resulted in 2,152, 1,261 and 1,132 nucleotide positions, respectively, for the mt, vWF and A2AB genes. The mt, vWF and A2AB data sets contain 810, 497 and 393 informative sites, respectively. Phylogenetic analyses (parsimony, minimum evolution with Tamura–Nei[27] and Logdet[28] distances, maximum likelihood under the HKY85 (ref. 29) model) were conducted with PAUP 4.0d52-54, written by D. L. Swofford. The mitochondrial tree was rooted using platypus and marsupial sequences. The vWF tree was rooted with the sloth[7]; alternatively, rooting with either hedgehog or rodents supports the 'African' clade and contradicts insectivore monophyly. For the A2AB tree, sequences with the suffix B are from the A2AB subfamily. GuineaPigA and GuineaPigC sequences are from other subfamilies in the α-2 adrenergic receptor family and were used as outgroups. Bootstrap analyses used full heuristic searches with 500 replications for parsimony and minimum

evolution and 100 replications for maximum likelihood. Shape parameters for the gamma distribution were estimated from minimum length trees[26] and were 0.32 (mtDNA), 0.59 (vWF) and 0.52 (A2AB).

Divergence times. 12S rRNA transversions accumulated linearly as far back as the eutherian–metatherian split[24]. Nine independent cladogenic events were selected based on 12S rRNA sequence availability and paleostratigraphic data[10,24,30] (for example, *Rattus* to *Mus* (14 Myr); *Sus* to *Tayassu* (45 Myr); ruminants to Cetacea (60 Myr); *Erinaceus* to Metatheria (130 Myr)). Relative rates were calculated in reference to xenarthrans. Tamura–Nei transversion distances (transversions only) were adjusted for relative rate differences[30] against the xenarthran standard. Rate-adjusted estimates of sequence divergence were regressed against paleostratigraphic divergence estimates for each of the nine calibration points (origin forced through zero; $r^2 = 0.97$; $P = 0.0000002$). The resulting equation (divergence time (in Myr) = sequence divergence/0.00063) was used to estimate interordinal divergence times after making similar adjustments for relative rates. Additional details will be presented elsewhere (M.S., manuscript in preparation).

Received 23 December 1996; accepted 18 April 1997.

1. Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356,** 121–125 (1992).
2. Huxley, T. H. On the application of the laws of evolution to the arrangement of the Vertebrata, and more particularly to the Mammalia. *Proc. R. Soc. Lond.* **43,** 649–662 (1880).
3. Matthew, W. D. The Carnivora and Insectivora of the Bridger Basin, Middle Eocene. *Mem. Am. Mus. Nat. Hist.* **9,** 291–567 (1909).
4. MacPhee, R. D. E. & Novacek, M. J. in *Mammal Phylogeny* Vol. 2, *Placentals* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 13–31 (Springer, New York, 1993),
5. Carroll, R. L. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
6. Gheerbrant, E., Sudre, J. & Cappetta, H. A Palaeocene proboscidean from Morocco. *Nature* **383,** 68–70 (1996).
7. Lavergne, A., Douzery, E., Stichler, T., Catzeflis, F. M. & Springer, M. S. Interordinal mammalian relationships: evidence for paenungulate monophyly is provided by complete mitochondrial 12S rRNA sequences. *Mol. Phyl. Evol.* **6,** 245–258 (1996).
8. Madsen, O., Deen, P. M. T., Pesole, G., Saccone, C. & de Jong, W. W. Molecular evolution of mammalian aquaporin-2: further evidence that elephant shrew and aardvark join the paenungulate clade. *Mol. Biol. Evol.* **14,** 363–371 (1997).
9. Porter, C. A., Goodman, M. & Stanhope, M. J. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gene. *Mol. Phys. Evol.* **5,** 89–101 (1996).
10. Stanhope, M. J. *et al.* Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. *J. Mol. Evol.* **43,** 83–92 (1996).
11. Cope, E. D. The condylarthra. *Am. Nat.* **18,** 790–805, 892–906 (1884).
12. Fischer, M. S. & Tassy, P. in *Mammal Phylogeny* Vol. 2, *Placentals* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 217–234 (Springer, New York, 1993).
13. de Jong, W. W., Zweers, A. & Goodman, M. Relationship of aardvark to elephants, hyraxes and sea cows from α-crystallin sequences. *Nature* **292,** 538–540 (1981).
14. de Jong, W. W., Leunissen, J. A. M. & Wistow, G. J. in *Mammal Phylogeny* Vol. 2, *Placentals* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 5–12 (Springer, New York, 1993).
15. Faith, D. P. Cladistic permutation tests for monophyly and nonmonophyly. *Syst. Zool.* **40,** 366–375 (1991).
16. Kishino, H. & Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29,** 170–179 (1989).
17. Glover, T. D. Aspects of sperm production in some east African mammals. *J. Reprod. Fertil.* **35,** 45–53 (1973).
18. Hartenberger, J. L. Hypothese paleontologique sur l'origine des Macroscelidea (Mammalia). *C.R. Acad. Sci.* **302,** 247–249 (1986).
19. Novacek, M. in *Macromolecular Sequences in Systematic and Evolutionary Biology* (ed. Goodman, M.) 3–41 (Plenum, New York, 1982).
20. Sibley, C. G. & Ahlquist, J. E. Reconstructing bird phylogeny by comparing DNAs. *Sci. Am.* **254,** 82–92 (1986).
21. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381,** 226–229 (1996).
22. Springer, M. S., Hollar, L. J. & Burk, A. Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. *Mol. Biol. Evol.* **12,** 1138–1150 (1995).
23. Thompson, J. D., Higgins, G. D. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).
24. Springer, M. S. & Douzery, D. Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. *J. Mol. Evol.* **43,** 357–373 (1996).
25. De Rijk, P., Van de Peer, Y., Chapelle, S. & De Wachter, R. *Nucleic Acids Res.* **22,** 3495–3501 (1994).
26. Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–514 (Sinauer, Sunderland, MA, 1996).
27. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10,** 512–526 (1993).
28. Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11,** 605–612 (1994).
29. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **21,** 160–174 (1985).
30. Arnason, U., Gullberg, A., Janke, A. & Xu, X. Pattern and timing of evolutionary divergences among hominoids based on analyses of complete mtDNAs. *J. Mol. Evol.* **43,** 650–661 (1996).

Acknowledgements. This work was supported by the Alfred P. Sloan Foundation, the European Commission, the NSF, the Nuffield Foundation and the Royal Society. We thank D. Willemsen for technical assistance, D. Swofford for permission to use PAUP 4.0d52-54, and F. Catzeflis, E. Harley, J. Kirsch, G. Olbricht, J. Wensing and the Noorder Zoo for tissue samples.

Chapter Four

Highly Congruent Molecular Support for a Diverse Superordinal Clade of Endemic African mammals

Michael J. Stanhope, Ole Madsen, Victor G. Waddell, Gregory C. Cleven, Wilfried W. de Jong and Mark S. Springer. 1998. Highly Congruent Molecular Support for a Diverse Superordinal Clade of Endemic African mammals. *Mol. Phylogenet. Evol.* **9**: 501-508

A solution to higher level mammalian phylogeny is going to depend on the congruent establishment of superordinal groupings followed by a linking together of these clades. We present congruent and convincing evidence from four disparate nuclear protein coding genes and from a tandem alignment of the 12S–16S mitochondrial region, for a superordinal clade of endemic African mammals that includes elephant shrews, aardvarks, golden mole, elephants, sirenians, and hyraxes. Because of strong support for golden mole as part of this clade, the Insectivora are rendered paraphyletic or polyphyletic, with constrained monophyly of the insectivores judged significantly worse in the vast majority of tests. Branching arrangement within this clade remains highly uncertain; however, a tandem alignment of the protein coding genes suggests elephant shrew is the earliest African lineage. None of the individual data sets or combinations of data sets support the widely held view of a mirorder Tethytheria (Sirenia/Proboscidea), although only a tandem alignment of protein coding and mitochondrial loci significantly rejects this association. The majority of the data sets and analyses provide strong support for Caviomorpha as part of a monophyletic Rodentia.

INTRODUCTION

Few areas of systematic investigation have attracted more interest than higher level mammalian phylogeny. The evolutionary history underlying the branching arrangement of the 18 orders of eutherian mammals has been an area of active research and controversy for over a century. The resolution of this complex problem undoubtedly lies through the stepwise construction of the true tree through a gradual assembling of superordinal groupings as they become established. Phylogenetic congruence between disparate data sets remains our best means of evaluating which branching arrangement most likely represents the true tree. We feel we have congruently established one such superordinal grouping that will have important implications in the way we regard mammalian evolution.

Aardvarks (order Tubulidentata), elephant shrews (order Macroscelidea), and the paenungulates (orders Sirenia, Proboscidea, and Hyracoidea) are mammals of presumed African origin that we have demonstrated, using a variety of different types of molecular loci, form a strongly supported monophyletic superordinal grouping. This same congruent cluster of five orders comes from amino acid sequence analysis of alphaA-lens crystallin (de Jong *et al.,* 1993), and DNA sequence analysis of exon 28 of the gene encoding von Willebrand Factor (vWF; Porter *et al.,* 1996), exon 1 of the interphotoreceptor retinoid binding protein gene (IRBP; Stanhope *et al.,* 1996), and aquaporin-2 (AQP2; Madsen *et al.,* 1997). Most recently we have identified an important new addition to this clade, in the golden mole (Springer *et al.,* 1997). This insectivore joins this superordinal group with the same convincing bootstrap support and does so at the exclusion of the other insectivores in our data sets (hedgehog, mole, and shrew), rendering the order Insectivora either paraphyletic or polyphyletic. This latest phylogenetic hypothesis of ours is congruently supported from DNA sequence analysis of the mitochondrial 12S + tRNA valine + 16S, exon 28 vWF, and the gene encoding alpha adrenergic receptor. These genes are of highly disparate function (arguably, in the case of the mtDNA ribosomal loci vs the nuclear protein encoding loci, as different as possible) and come from different genomes. This therefore adds several further convincing loci to the growing list of genes in support of this African clade and provides an important addition to this phylogenetic grouping, in the African golden mole. Importantly, like the previous five orders of mammals, the golden moles are endemic African mammals.

This clade, now composed of five orders and a portion of a sixth, has important implications for the study of mammalian evolution. The monophyly of this superordinal grouping of endemic African mammals implies that there was an extensive African radiation from a single common ancestor that gave rise to ecologically divergent adaptive types. Geographic evidence, in other words, adds to the molecular data in support of the African origin of this clade. This adaptive radiation of the African clade parallels endemic radiations of other vertebrate taxa on southern hemisphere continents during the break-up of Gondwanaland. Notably, the divergence times for members of this clade, estimated from 12S rRNA transversions, suggest that the base of this radiation occurred during Africa's window of isolation in the Cretaceous before tenuous land connections were developed with Europe in the early Cenozoic, which in turn implies that the conventional view of a European condylarth ancestry for the origin of these African mammals is incorrect [see Springer *et al.* (1997) for a more complete discussion]. Lastly, the demonstration of a paraphyletic or polyphyletic Insectivora calls for a revaluation of our current taxonomic arrangement for Eutheria.

Any phylogenetic hypothesis, particularly one with these sorts of implications, needs to be examined critically and thoroughly. The purpose of the present paper is to expand several of these data sets, in terms of taxonomic representation (including the addition of the golden mole to the AQP2 data set where it had previously been missing), and to examine in more detail the level and nature of support for this African phylogenetic hypothesis, as well as the branching arrangement within the clade.

MATERIALS AND METHODS

This study expands, in terms of taxonomic representation, several already existing higher level mammalian data sets. The previous IRBP (interphotoreceptor retinoid binding protein) data set (Stanhope *et al.,* 1996) was rooted at sloth, which may not necessarily represent the earliest eutherian branch. For the present analysis we use the opossum sequence from Stanhope *et al.* (1992) as an outgroup. The other addition to this IRBP data set includes our cetacean sequences from Smith *et al.* (1996), for a total of 31 species, representing all orders except Monotremata and Pholidota (see Table 1 for a complete list of sequences and their accession numbers).

Additions to the vWF (von Willebrand Factor) data set of Springer *et al.* (1997) include human (*Homo sapiens*), spider monkey (*Ateles belzebuth*), galago (*Galago crassicaudatus*), rat (*Rattus norvegicus*), mouse (*Mus domesticus), Tadarida brasiliensis, Cynopterus sphinx*, dog (*Canis familiaris*), and harbor porpoise (*Phocoena phocoena*). Of these, human, galago, dog, mouse, *Tadarida,* and *Cynopterus* appeared in the earlier trees of Porter *et al.* (1996). Additional sequence positions were added to several previously incomplete sequences from Porter *et al.* (1996), including cow, galago, *Tadarida,* tree shrew, and cat, completing (or very nearly) the 1.3-kb exon 28 fragment for these taxa. Harbor porpoise and spider monkey are newly acquired and no portion of these sequences appears elsewhere. PCR and sequencing for vWF is as described in Porter *et al.* (1996). This particular vWF data set included 32 species, representing all orders of eutherian mammals (Table 1). Analyses were conducted with trees rooted at sloth.

To the previous data set for alpha-2B adrenergic receptor (A2AB) of 11 taxa (Springer *et al.,* 1997), we add five new sequences including the opossum (*Didelphis virginiana*), rabbit (*Oryctolagus cuniculus*), cow (*Bos taurus*), horse (*Equus caballus*), and dugong (*Dugong dugon*). PCR and sequencing are as described in Springer *et al.* (1997). The A2AB primers amplify an approximately 1.2-kb fragment; however, not all these data can be unambiguously aligned because of a glutamic acid repeat approximately 850 bp downstream of the start codon of this single exon. This repeat region, as well as a 9-bp stretch of ambiguous indels approximately 750 bp into the exon, were removed from the alignment, resulting in 1137 aligned positions for 16 taxa representing 12 orders (Table 1). The opossum, however, could not be unambiguously aligned between 626 and 701 and between 775 and 890 of this 1137-bp sequence alignment. As a consequence, in analyses of these A2AB data with opossum as outgroup, these ambiguous regions were excluded from the sequence alignment. Because of the undesirability of excluding approximately 20% of the data, A2AB was also analyzed without opossum and with rodents instead serving as the outgroup.

The original aquaporin-2 (AQP2) data set included 12 mammalian species (Madsen *et al.,* 1997). To this sequence alignment we add five additional species, including the guinea pig (*Cavia porcellus*), dugong (*D. dugon*), hedgehog (*Erinaceus europaeus*), mole (*Talpa europaea*), and golden mole (*Amblysomus hottentotus*), for a total data set of 17 taxa representing 14 orders for the aligned 328 bp exon 1 (Table 1). PCR and sequencing are as described in Madsen *et al.* (1997). All trees were rooted at armadillo.

The previous 12S–16S mammalian data set included 25 taxa, representing 12 placental orders, marsupials, and monotremes (Springer *et al.,* 1997). To this data set we add African elephant, *Loxodonta africana* (Table 1). All potentially ambiguous regions of this alignment were exclude from analyses resulting in 2146 sequence positions. PCR and sequencing are as described in Springer *et al.* (1997). Trees were rooted at platypus.

All of the foregoing sequence alignments are available from the EMBL file server under the following

TABLE 1

Species, Genes, and the Corresponding GenBank Accession Numbers of Sequences Used in This Study

Species	IRBP	vWF	A2AB	AQP2	12S–16S
Three toed sloth, *Bradypus tridactylus*	U48708	U31603			
Armadillo, *Dasypus novemcinctus*				Y10637	
Elephant shrew, *Elephantulus rufescens*	U48584	U31612			U97339
Elephant shrew, *Macroscelides proboscideus*			Y12524	Y10630	
Dugong, *Dugong dugon*	U48583	U31608	Y15946*	Y15948*	
Manatee, *Trichechus manatus*					U97337, U61083
Aardvark, *Orycteropus afer*	U48712	U31617	Y12522	Y10632	U97338
Hyrax, *Procavia capensis*	U48586	U31619	Y12523	Y10631	U97335, U61084
African elephant, *Loxodonta africana*	U48711	U31615			U60182, AF039436*
Asian elephant, *Elephas maximus*		U31611	Y12525	Y10629	
Golden mole, *Amblysomus hottentotus*		U97534	Y12526	Y15952*	U97336, M95108
Rough-toothed dolphin, *Steno bredanensis*	U48713				
Pilot whale, *Globicephala macrorhynchus*	U50821				
Harbor porpoise, *Phocoena phocoena*		AF061060*			
Gray whale, *Eschrichtius robustus*	U50649				
Minke whale, *Balaenoptera acutorostrata*	U50820				
Fin whale, *Balaenoptera physalus*					X61145
Blue whale, *Balaenoptera musculus*					X72204
Giant sperm whale, *Physeter catodon*	U50818				
Pygmy sperm whale, *Kogia breviceps*	U50819				
Pig, *Sus scrofa*	U48588	S78431			
Cow, *Bos taurus*	J04441	X63820, AF004285*	Y15944*	Y10633	J01394
Horse, *Equus caballus* (vWF: *E. asinus*)	U48710	U31610	Y15945*	Y10634	X79547
White rhino, *Ceratotherium simum*		U31604			
Mole rat, *Spalax polinicus*	U48589	U31621			
Mouse, *Mus domesticus*	Z11813	U27810	L00974		J01420
Rat, *Rattus norvegicus*		U50044	M32061	D13906	X14848
Guinea pig, *Cavia porcellus*			U25723	Y15947*	L35585
Agouti, *Dasyprocta agouti*		U31607			
Rabbit, *Oryctolagus cuniculus*	Z11812	U31618	Y16189*	Y10636	
Pangolin, *Manis* sp.		U97535		Y10635	U97340, U61079
Shrew, *Sorex palustris*	U48587				U97343
Hedgehog, *Erinaceus europaeus*		U97536	Y12521	Y15949*	X88898
Mole, *Talpa europaea*			Y12520	Y15951*	
Macrotus californicus	U48585				
Tonatia bidens	Z11810				
Tonatia silvicola	Z11829				
Megaderma lyra		U31616			
Tadarida brasiliensis		U31623, AF061061*			
Pteropus hypomelanus	Z11809				
Cynopterus sphinx	U48709	U31605			
Dobsonia moluccensis		U31609			
Cat, *Felis catus*	Z11811	U31613, AF061062*			U20753
Dog, *Canis familiaris*		L76227		Y10638	
Harbor seal, *Phoca vitulina*					X63726
Gray seal, *Halichoerus grypus*					X72004
Flying lemur, *Cynocephalus variegatus*	Z11807	U31606			
Tree shrew, *Tupaia glis*	Z11808	U31624, AF061063*			
Galago, *Galago crassicaudatus*	Z11805	U31614, AF061064*			
Tarsier, *Tarsius syrichta*	Z11806				
Spider monkey, *Ateles belzebuth*		AF061059*			
Orangutan, *Pongo pygmaeus*					D38115
Gorilla, *Gorilla gorilla*					D38114
Chimpanzee, *Pan troglodytes*		U31620			
Human, *Homo sapiens*	J05253	M25851	M34041	D31846	J01415
Opossum, *Didelphis virginiana*	Z11814		Y15943*		Z29573
Bandicoot, *Echymipera kalubu*					U97342
Monito del monte, *Dromiciops gliroides*					U97341, U61073
Platypus, *Ornithorhynchus anatinus*					U33498, X83427

Note. Asterisks indicate a sequence new to this study. Common names for bats tend to be highly ambiguous and exhibit considerable geographic variability, and therefore are not listed.

accession numbers: IRBP, DS33922; vWF, DS33921; A2AB, DS33923; AQP2, DS33920; 12S–16S, DS33899.

Data were analyzed by neighbor joining (NJ), maximum parsimony (MP), and maximum likelihood (DNAML) methods using PHYLIP (Felsenstein, 1993). Maximum likelihood analyses of amino acid sequences were conducted using PUZZLE (Strimmer and von Haeseler, 1996). The robustness of the phylogenetic hypotheses were examined by bootstrapping (Felsenstein, 1985), and Bremer decay indexes (Bremer, 1994; using PAUP 4.0d52–54, written by D. L. Swofford). For the NJ and MP analyses, bootstrapping involved 1000 replicates and for the DNAML analyses 100 replicates. The DNAML bootstrap analyses for the larger data sets were performed on a Silicon Graphics–Crimson computer. Most parsimonious trees were determined by randomizing the input order 50 times. Maximum likelihood analyses were performed with the global branch swapping option and an expected transition/transversion ratio of 2.0. Neighbor joining analyses were performed using Kimura two-parameter divergence matrices and a 2.0 transition/transversion ratio. In several instances significance tests were performed between user defined or constrained trees and the maximum parsimony tree as well as the maximum likelihood tree. The parsimony tests were conducted using the method proposed by Templeton (1983), and the maximum likelihood tests used the method of Kishino and Hasegawa (1989). The null hypothesis was rejected at $P < 0.05$ in both these tests.

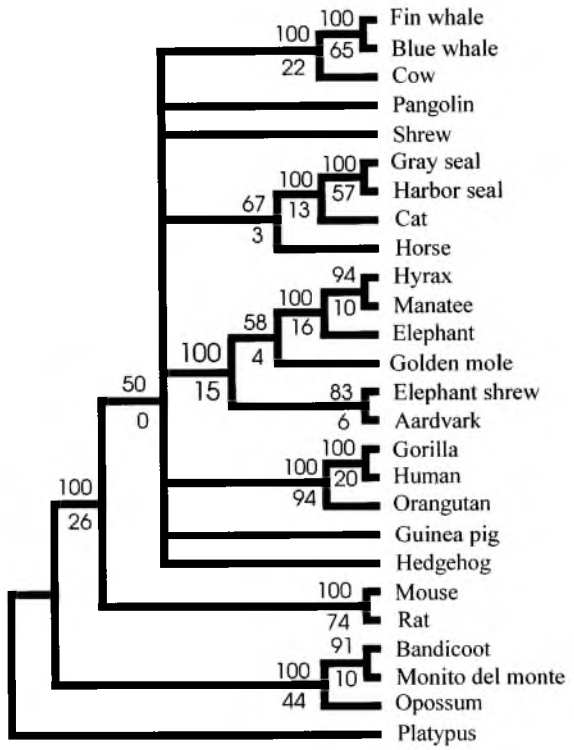


FIG. 1. Maximum parsimony bootstrap tree for the 12S–16S data set. Numbers above the branch indicate percentage bootstrap figures from 1000 replications of the data; numbers below the branch refer to the Bremer decay index. Nodes receiving less than 50% bootstrap support are collapsed (4141 steps; CI = 0.421; RI = 0.462).

RESULTS AND DISCUSSION

Phylogenetic analyses of these various data sets all result in an African clade composed of elephant shrews, aardvark, the paenungulates, and the golden mole, although the branching arrangement within this clade does not always agree amongst the various loci (Figs. 1–4). The addition of the elephant to the 12S–16S data set has resulted in the traditional Paenungulata, however, with a hyrax/manatee clade rather than the expected Sirenia/Proboscidea grouping (Fig. 1). Constraining the data to support a manatee/elephant clade added 7 substitutions to the most parsimonious tree and was judged not significantly worse than the MP or highest likelihood tree. As in many other mitochondrial data sets the caviomorph rodent, here represented by the guinea pig, does not group with the myomorphs, although it only costs 2 additional steps to support a monophyletic Rodentia, a change which was judged not significantly different from the MP or highest likelihood tree. The branching arrangement for the African clade has the golden mole joining the paenungulates followed by an elephant shrew/aardvark clade. Bootstrap support for this superordinal African clade was in excess of 93% for all types of analyses (Table 2) involving this locus, with a decay index of 15 (Fig. 1). Insectivores are paraphyletic in all 12S–16S analyses; golden mole joins the Africans, with shrew and hedgehog separated. It costs, however, a mere 5 additional steps to support a grouping of shrew and hedgehog, which was judged not significantly different from the MP tree or the DNAML tree. In sharp contrast, a completely monophyletic Insectivora costs an additional 48 steps to the MP tree and was judged significantly different from both the MP and DNAML tree.

IRBP also supports this same African clade; however, the sequence of golden mole for this gene is still unavailable, because of persistent difficulties in amplification. This tree differs from our earlier analyses (Stanhope *et al.*, 1996) in that it is now rooted with the opossum sequence (Stanhope *et al.*, 1992), and that it includes a range of cetaceans (Smith *et al.*, 1996). This more dense phylogenetic data set, rooted at marsupials, has done nothing to diminish the support for the African clade with 100% bootstrap support for this grouping in all types of analyses (Table 2), and a decay index of 38. This level of Bremer support is double that which defines Cetacea/Artiodactyla, another superordinal grouping which receives 100% bootstrap support, and for which there is now no argument. The African clade in this data set consists of a strongly supported Paenungulata, joined by an aardvark/elephant shrew

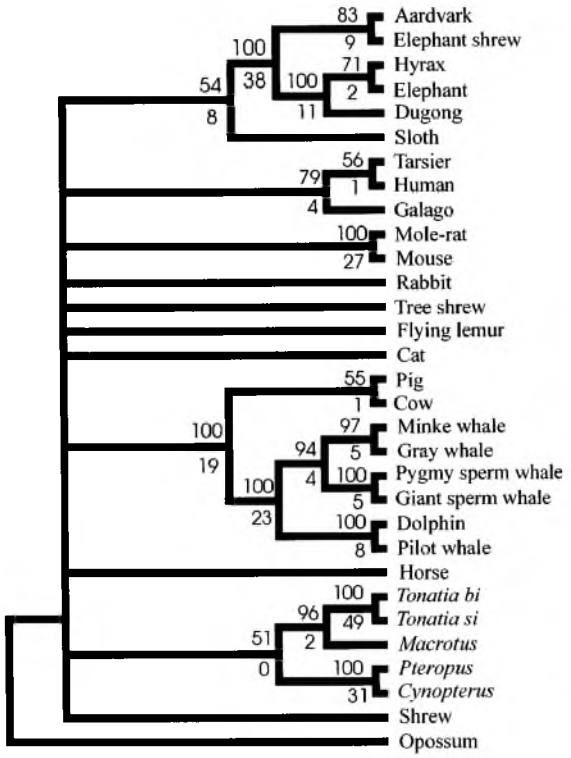


FIG. 2. Maximum parsimony bootstrap tree for the IRBP data set. Numbers above the branch indicate percentage bootstrap figures from 1000 replications of the data; numbers below the branch refer to the Bremer decay index. Nodes receiving less than 50% bootstrap support are collapsed (2922 steps; CI = 0.483; RI = 0.495). *Tonatia bi, Tonatia bidens; Tonatia si, Tonatia silvicola.*

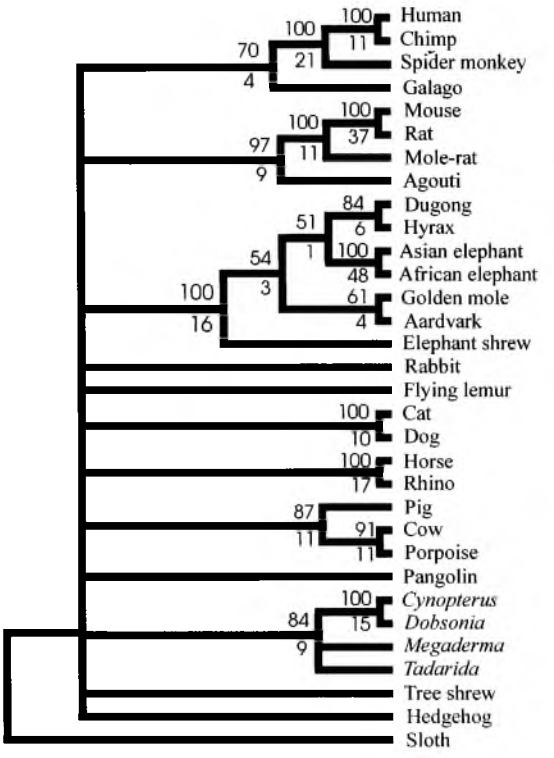


FIG. 3. Maximum parsimony bootstrap tree for the vWF data set. Numbers above the branch indicate percentage bootstrap figures from 1000 replications of the data; numbers below the branch refer to the Bremer decay index. Nodes receiving less than 50% bootstrap support are collapsed (3088 steps; CI = 0.444; RI = 0.434).

clade. Tethytheria receives no support in any of the analyses and it requires 7 additional substitutions to support this association; however, this constrained topology was judged not to be significantly different from both the MP and DNAML trees. Edentates, represented by the sloth, do not join at the base of Eutheria, and instead join the African clade, with that overall grouping in an unstable position well within the other orders. The first eutherian branch is shrew.

As in the other loci, the addition of several taxa to the vWF data set (Fig. 3) has done nothing to upset the strong support for the African clade. The clade receives 100% (or very nearly) bootstrap support in all types of analyses (Table 2). Alternative rootings (rodents and primates, respectively) did nothing to diminish support for this grouping. The vWF data have a golden mole/aardvark clade joining the Paenungulata, followed by the elephant shrew. As in the previous two data sets, Tethytheria receives no support and it requires 7 additional steps to support an elephant/dugong clade, an arrangement which was judged not significantly different from either the MP or DNAML trees. With the inclusion of the cetacean this data set now has a representative from all orders of Eutheria and typical of nearly all other types of data, this cetacean joins the

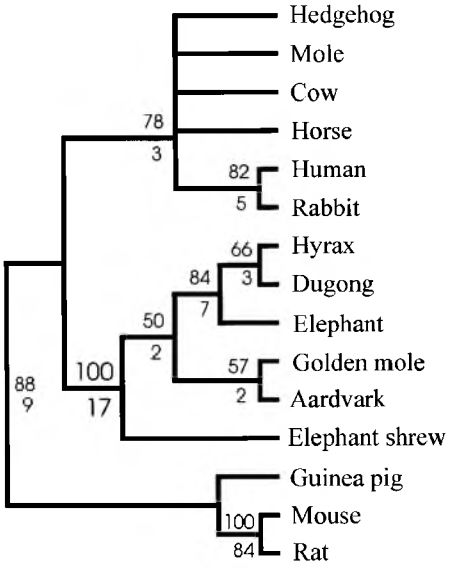


FIG. 4. Maximum parsimony bootstrap tree for the A2AB data set. Numbers above the branch indicate percentage bootstrap figures from 1000 replications of the data; numbers below the branch refer to the Bremer decay index. Nodes receiving less than 50% bootstrap support are collapsed (1192 steps; CI = 0.664; RI = 0.491).

TABLE 2

Bootstrap Support for the African Clade, from Various Loci and Combinations of Loci, Using Different Phylogenetic Methods

	MP	TVP	DNAML	NJ
IRBP	100	100	100	100
vWF	100	98	100	100
A2AB	100	99	97	99
AQP2	—	NA	54	56
12S–16S	100	93	100	97
vWF + A2AB + AQP2	100	100	100	100
vWF + A2AB + AQP2 + 12S–16S	100	100	100	100

Note. MP, maximum parsimony; TVP, transversion parsimony; DNAML, maximum likelihood; NJ, neighbor joining; NA, not applicable. A2AB values come from analyses with rodents as the outgroup.

artiodactyls with convincing bootstrap support. The caviomorph rodent in this data set, represented by the agouti, joins the myomorph rodents with 96–98% bootstrap support, and a decay index of 9. The best tree that supports the monophyly of the Insectivora has a hedgehog/golden mole clade joining aardvark, still within the African clade, and adds 37 substitutions, making it significantly worse than the MP tree. Insectivora monophyly was also judged significantly worse using the likelihood test.

The A2AB data set, rooted at rodents (Fig. 4), and with the inclusion of several new taxa, supports the African clade with bootstrap support near 100% for all types of analyses (Table 2) and a decay index of 17. Alternatively, rooting the tree with the opossum, after removal of the 190 bp of ambiguous sequence alignment, still results in the African superordinal grouping, although with somewhat less bootstrap support (MP: 64%; NJ: 75%; DNAML: does not support the grouping). The addition of the dugong completes the ordinal representation for the Paenungulata, but again there is no support for Tethytheria, with dugong joining hyrax instead of the elephant. A constrained dugong/elephant clade adds 3 substitutions to the MP tree and was judged not significantly different from both the most parsimonious and highest likelihood trees. All types of analyses of this A2AB data set result in an African clade composed of the Paenungulata joined by an aardvark/golden mole clade, followed by the elephant shrew. The two other insectivores, mole and hedgehog, join together in the MP and DNAML trees, albeit not with convincing bootstrap support. Forcing the monophyly of Insectivora results in hedgehog/mole/golden mole joining the remaining Africans (i.e., elephant shrew/aardvark/Paenungulata), with an additional 40 substitutions to the MP tree, which was judged significantly worse than that MP tree. Likelihood tests yielded a similar result. The monophyly of Rodentia is strongly supported regardless of how the tree is rooted. For example, trees rooted at human result in bootstrap support of 90, 79, and 93 for MP, NJ, and DNAML analyses, respectively. Alternatively, rooting the tree at cow results in bootstrap support of 91, 80, and 89 for the same three tests.

The addition of several taxa to the AQP2 data, including the golden mole, results in a set of 13 most parsimonious trees (369 steps; CI = 0.570; RI = 0.426) all of which support the monophyly of a clade consisting of the paenungulates, aardvark, golden mole, elephant shrew, and hedgehog. All MP trees have an association of elephant shrew and hedgehog joining the rest of the Africans. NJ and DNAML trees on the other hand support the monophyly of the Africans at the exclusion of the hedgehog. Analyzing these data at the amino acid sequence level using parsimony, neighbor joining, and maximum likelihood methods also supports the monophyly of the Africans. Importantly, within this 108-residue alignment, there are three positions where the same amino acid replacement occurs solely in all the Africans, including the golden mole (15 A → S, 58 A → T, 107 I → L). Constraining this data set to support the monophyly of Insectivora adds 7 substitutions and was judged not to be significantly worse than the minimum length tree, or the highest likelihood tree.

One of the important controversies in modern systematics is the debate centered around whether multiple data sets should be combined in phylogenetic analysis [see Miyamoto and Fitch (1995) for a review]. There are both extreme and intermediate views on this subject, with the intermediate position suggesting that data sets should only be combined when they do not strongly support conflicting trees (de Queiroz, 1993). Our data sets do not comprise a set of strongly supported conflicting topologies and thus combining them is arguably appropriate. Combining the vWF, A2AB, and AQP2 data into a concatenated alignment of protein encoding genes involving 14 taxa (human, rabbit, elephant shrew, elephant, dugong, hyrax, aardvark, golden mole, mouse, rat, guinea pig, hedgehog, cow, and horse; 2722 bp), using all positions, rooted at rodents, results not surprisingly in 100% support for the African clade (regardless of method of analysis). The branching arrangement was in agreement for all types of analyses (MP, NJ, DNAML), and consisted of the Paenungulata (90, 99, 97% bootstrap support), composed of a hyrax/dugong clade (77, 72, 75% support), joined by elephant, followed next by a golden mole/aardvark clade (84, 65, 78%). This collection of five taxa receives 88, 100, and 90% support, respectively, with the elephant shrew the last to join that group. A constrained tree that supports Tethytheria adds 7 substitutions to the MP tree and was judged not significantly worse than both the MP and highest likelihood tree. One other feature of interest in this tree is that the caviomorph rodent (represented by a hybrid guinea pig/agouti sequence) joins the two myomorph rodents with 96–100% bootstrap

support, regardless of how the tree is rooted or method of analysis (i.e., MP, NJ, and DNAML analyses were also performed with human and cow as outgroups). Analyzing these data at the amino acid sequence level yields very similar results. Adding IRBP to this concatenated protein encoding data set, now with the exclusion of golden mole, since it is not available for IRBP (10 taxa: human, mouse, elephant, dugong, hyrax, aardvark, elephant shrew, cow, horse, and rabbit; 3972 bp), does little to change the overall picture. One feature of interest in this data set is increased support for dugong/hyrax, with a constrained Tethytheria now adding a total of 16 steps to the MP tree, a change which is still, however, judged not significantly different from the MP or highest likelihood tree. A combination of A2AB and AQP2 (15 taxa: human, rabbit, mole, hedgehog, rat, mouse, guinea pig, elephant, dugong, hyrax, aardvark, golden mole, elephant shrew, cow, and horse; 1473 bp) provides a nuclear gene perspective on the monophyly of two non-African insectivore families (i.e., hedgehog and mole). Parsimony and maximum likelihood support the grouping, albeit not with convincing bootstrap support, while neighbor joining does not. Constrained monophyly of Insectivora (i.e., golden mole, hedgehog, and mole) in this A2AB + AQP2 data set adds 44 substitutions and was judged significantly worse than both the MP and DNAML trees.

An alignment of vWF, A2AB, AQP2, and the 12S–16S data (4856 bp) for all of the same taxa as vWF + A2AB + AQP2 with the exception of rabbit and with the caveat that sirenians are represented by a hybrid manatee–dugong sequence, results in trees (MP, NJ, and DNAML) each of which have a different branching arrangement for the African clade. The differences lie in regard to which taxon or clade represents the first branch. All analyses support the African clade at 100%. Paenungulata receives 96, 100, and 98% support, respectively. Dugong and hyrax join within the Paenugulata with 89, 91, and 92% support. The branching arrangement of the aardvark, golden mole, and elephant shrew is unresolved. A constrained tree that supports Tethytheria adds 21 substitutions to the MP tree and was judged significantly worse than both the MP and highest likelihood tree. As in most of the preceding data sets the caviomorph rodent joins the two myomorph rodents with convincing bootstrap support (96–100%), regardless of how the tree is rooted or method of analysis. Another possible data set combination includes all nuclear genes (i.e., including IRBP) and the mitochondrial loci (nine species for 5996 bp) at the exclusion of golden mole because of its absence in IRBP. This results in 80–97% bootstrap support for hyrax/dugong, with trees supporting Tethytheria significantly worse than the MP and highest likelihood trees, and requiring 22 additional steps.

One of the emerging curiosities in the molecules versus morphology debate, at least in regard to higher level mammalian phylogeny, is the relative paucity of convincing molecular support for Tethytheria compared to the wealth of morphological evidence in support of this clade. Shoshani (1992), for example, lists 10 characters which he feels are likely synapomorphies for a Sirenia/Proboscidea clade, compared to 13 characters in support of Paenungulata. Lavergne *et al.* (1996) in an interordinal examination of the 12S gene present a rare case of molecular support for Tethytheria. Ozawa *et al.* (1997) in an examination of the evolutionary history of the extinct woolly mammoth and Stellar's sea cow, using the cytochrome *b* gene also present some support for Tethytheria, although bootstrap values are very minimal, and several analyses (parsimony, for example) do not support the clade at all. The majority of other molecular data sets have not supported this grouping (see for example, Stanhope *et al.,* 1993; de Jong *et al.,* 1993; Graur *et al.,* 1997), and we now add five additional loci to this growing list. The tandem alignments presented here represent the largest DNA sequence data sets pertinent to the problem; interestingly, Tethytheria is deemed gradually less likely with increasing data set size, to the point where it is eventually judged significantly worse than both the MP and highest likelihood trees. One of the arguments of combining data sets is that true (but weak) phylogenetic signal, masked by homoplasy, can be amplified above the phylogenetic noise, to the point where it becomes detectable (Barrett *et al.,* 1991). Whether the present situation is a case in point, or whether it is more a case of random phylogenetic noise, with increasing data set size, eventually acquiring an incorrect directionality (see for example Naylor and Brown, 1997), remains to be seen. This is, in our opinion, an issue which is central to modern molecular systematics because of the ever increasing amount of data which will be analyzed in tandem alignments.

All five of these phylogenetic loci, of highly disparate functions, support a clade of endemic African mammals, generally with highly convincing bootstrap support. In fact there are no published molecular data sets at the DNA sequence level which include the members of this clade that do not support this grouping. Congruent support of this nature is an extreme rarity in higher level systematics of mammals. The only exception to this is the close evolutionary relationship of Cetacea and Artiodactyla, for which there can now be no doubt. We submit that the African clade is certainly comparable in support; however, the precise nature of the branching arrangement within it remains highly uncertain, and indeed it remains a distinct possibility that there are additional members. We also feel these data indicate that there can be little doubt that Insectivora is not a natural monophyletic assemblage; however, the extent of insectivore paraphyly (or polyphyly) and the possible reclassification of the group will have to await

additional taxon sampling to include all the current families of Insectivora.

ACKNOWLEDGMENTS

We thank the Nuffield Foundation, Royal Society, Alfred P. Sloan Foundation, European Commission, and the National Science Foundation for support, Diederik Willemsen for technical assistance, Dr. Calvin Porter for his contribution of sequence data, and Drs. F. Catzeflis, E. Harley, J. Kirsch, G. Olbricht, J. Wensing, and the Noorder Zoo for tissue samples.

REFERENCES

Barrett, M., Donoghue, M. J., and Sober, E. (1991). Against consensus. *Syst. Zool.* **40:** 486–493.

Bremer, K. (1994). Branch support and tree stability. *Cladistics* **10:** 295–304.

de Jong, W. W., Leunissen, J. A. M., and Wistow, G. J. (1993). Eye lens crystallins and the phylogeny of placental orders: Evidence for a Macroscelid-Paenungulate clade? *In* "Mammal Phylogeny: Placentals" (F. S. Szalay, M. J. Novacek, and M. C. McKenna, Eds.), pp. 5–12. Springer-Verlag, New York.

de Queiroz, A. (1993). For consensus (sometimes). *Syst. Biol.* **42:** 368–372.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39:** 783–791.

Felsenstein, J. (1993). PHYLIP (phylogeny inference package), Version 3.5c. Department of Genetics, University of Washington, Seattle. [Distributed by the author]

Graur, D., Gouy, M., and Duret, L. (1997). Evolutionary affinities of the order Perissodactyla and the phylogenetic status of the superordinal taxa Ungulata and Altungulata. *Mol. Phylogenet. Evol.* **7:** 195–200.

Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of the Hominoidea. *J. Mol. Evol.* **29:** 170–179.

Lavergne, A., Douzery, E., Stichler, T., Catzeflis, F. M., and Springer, M. S. (1996). Interordinal mammalian relationships: Evidence for paenungulate monophyly is provided by complete mitochondrial 12S rRNA sequences. *Mol. Phylogenet. Evol.* **6:** 245–258.

Madsen, O., Deen, P. M. T., Pesole, G., Saccone, C., and de Jong, W. W. (1997). Molecular evolution of mammalian aquaporin-2: Further evidence that elephant shrew and aardvark join the Paenungulate clade. *Mol. Biol. Evol.* **14:** 363–371.

Miyamoto, M. M., and Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* **44:** 64–76.

Naylor, G. J. P., and Brown, W. M. (1997). Structural biology and phylogenetic estimation. *Nature* **388:** 527–528.

Ozawa, T., Hayashi, S., and Mikhelson, V. M. (1997). Phylogenetic position of Mammoth and Stellar's sea cow within Tethytheria demonstrated by mitochondrial DNA sequences. *J. Mol. Evol.* **44:** 406–413.

Porter, C. A., Goodman, M., and Stanhope, M. J. (1996). Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand Factor gene. *Mol. Phylogenet. Evol.* **5:** 89–101.

Shoshani, J. (1992). The controversy continues: An overview of evidence for Hyracoidea-Tethytheria affinity. *Isr. J. Zool.* **38:** 233–244.

Smith, M. R., Shivji, M. S., Waddell, V. G., and Stanhope, M. J. (1996). Phylogenetic evidence from the IRBP gene for the paraphyly of toothed whales, with mixed support for Cetacea as a suborder of Artiodactyla. *Mol. Biol. Evol.* **13:** 918–922.

Springer, M. S., Cleven, G. C., Madsen, O., de Jong, W. W., Waddell, V. G., Amrine, H. M., and Stanhope, M. J. (1997). Endemic African mammals shake the phylogenetic tree. *Nature* **388:** 61–64.

Stanhope, M. J., Czelusniak, J., Si, J.-S., Nickerson, J., and Goodman, M. (1992). A molecular perspective on mammalian evolution from the gene encoding Interphotoreceptor Retinoid Binding Protein, with convincing evidence for bat monophyly. *Mol. Phylogenet. Evol.* **1:** 148–160.

Stanhope, M. J., Bailey, W. J., Czelusniak, J., Goodman, M., Si, J.-S., Nickerson, J., Sgouros, J. G., Singer, G. A. M., and Kleinschmidt, T. K. (1993). A molecular view of primate supraordinal relationships from the analysis of both nucleotide and amino acid sequences. *In* "Primates and Their Relatives in Phylogenetic Perspective" (R. D. E. MacPhee, Ed.), pp. 251–292. Plenum, New York.

Stanhope, M. J., Smith, M. R., Waddell, V. G., Porter, C. A., Shivji, M. S., and Goodman, M. (1996). Mammalian evolution and the Interphotoreceptor Retinoid Binding Protein (IRBP) gene: Convincing evidence for several superordinal clades. *J. Mol. Evol.* **43:** 83–92.

Strimmer, K., and von Haeseler, A. (1996). Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13:** 964–969.

Templeton, A. R. (1983). Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37:** 221–244.

Chapter Five

Molecular evidence for multiple origins of Insectivor and for a new order of endemic African insectivore mammals

Michael J. Stanhope, Victor G. Waddell, Ole Madsen, Wilfried W. de Jong, S. Blair Hedges, Gregory C. Cleven, Diana Kao and Mark S. Springer. 1998. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl. Acad. Sci. USA* **95**: 9967-9972

ABSTRACT **The traditional views regarding the mammalian order Insectivora are that the group descended from a single common ancestor and that it is comprised of the following families: Soricidae (shrews), Tenrecidae (tenrecs), Solenodontidae (solenodons), Talpidae (moles), Erinaceidae (hedgehogs and gymnures), and Chrysochloridae (golden moles). Here we present a molecular analysis that includes representatives of all six families of insectivores, as well as 37 other taxa representing marsupials, monotremes, and all but two orders of placental mammals. These data come from complete sequences of the mitochondrial 12S rRNA, tRNA-Valine, and 16S rRNA genes (2.6 kb). A wide range of different methods of phylogenetic analysis groups the tenrecs and golden moles (both endemic to Africa) in an all-African superordinal clade comprised of elephants, sirenians, hyracoids, aardvark, and elephant shrews, to the exclusion of the other four remaining families of insectivores. Statistical analyses reject the idea of a monophyletic Insectivora as well as traditional concepts of the insectivore suborder Soricomorpha. These findings are supported by sequence analyses of several nuclear genes presented here: vWF, A2AB, and α-β hemoglobin. These results require that the order Insectivora be partitioned and that the two African families (golden moles and tenrecs) be placed in a new order. The African superordinal clade now includes six orders of placental mammals.**

Throughout most of this century, the placental (eutherian) mammals with extant representation have been classified into 18 orders. During this period, the order Insectivora has been among the least stable higher taxa in Eutheria, both in terms of phylogenetic position and taxonomic content. Beginning with Huxley (1) and later embellished by Mathew (2), insectivores have been thought to possess features that rendered them closer to the ancestral stock of mammals. Despite this presumed central position of insectivores in the evolutionary history of mammals, the composition of the group never has been widely agreed on. The prevalent morphological view (3) suggests that the following extant families of "insectivores" descended from a single common ancestor and as such should be those groups that are regarded as the constituents of the order Insectivora (Lipotyphla): Soricidae (shrews), Tenrecidae (tenrecs), Solenodontidae (solenodons), Talpidae (moles), Erinaceidae (hedgehogs and gymnures), and Chrysochloridae (golden moles).

Butler (4) listed six morphological characteristics that, in his opinion, supported a monophyletic Insectivora including (*i*) absence of cecum; (*ii*) reduction of pubic symphysis; (*iii*) maxillary expansion within orbit, displacing palatine; (*iv*) mobile proboscis; (*v*) reduction of jugal; and (*vi*) hemochorial placenta. More recently, MacPhee and Novacek (3) have reviewed the evidence and concluded that characteristics (*i*) and (*ii*) support lipotyphlan monophyly, characteristic (*iii*) possibly does, and (*iv–vi*), as currently defined, do not, leaving two to three characteristics that, in their opinion, support the order Insectivora.

The six families of insectivores are most often grouped into two clades of subordinal rank: the Erinaceomorpha (hedgehogs) and the Soricomorpha (all other families). Within the Soricomorpha, Butler (4) suggested that the golden moles and tenrecs form a clade and that moles and shrews cluster together, followed by solenodons. MacPhee and Novacek (3), however, proposed three clades of subordinal rank: Chrysochloromorpha (Chrysochloridae), Erinaceomorpha (Erinaceidae), and Soricomorpha (Soricidae, Talpidae, Solenodontidae, and Tenrecidae). This latter organization is based on their view that the Chrysochloridae is "spectacularly autapomorphic." In their opinion, golden moles show no shared derived traits with the soricomorphs and therefore should be separated from that suborder. This recommendation echoes earlier views regarding the group that have suggested that golden moles are a separate order or suborder (5–7).

A recent molecular study of mammalian phylogeny, which included three insectivore families, demonstrated that golden moles are not part of the Insectivora but instead belong to a clade of endemic African mammals that also includes elephants, hyraxes, sea cows, aardvarks, and elephant shrews (8). Evidence for this now comes from a wide range of disparate molecular loci including the nuclear AQP2, vWF, and A2AB genes as well as the mitochondrial 12S–16S rRNA genes (8, 9). The fossil record of golden moles indicates that the geographic distribution of this group has been restricted to Africa throughout its temporal range (10, 11). The fossil record of tenrecs also suggests an African origin (10, 11). This paleontological record, along with the morphological study of Butler (4), suggests a possible common ancestry for golden moles and tenrecs. However, it also may be that the autapomorphic qualities of golden moles reflect their evolutionary history as a singular distinct lineage of African insectivores, separate from the rest of the order. At present, there is no published molecular phylogenetic perspective on the evolutionary history of the Tenrecidae, there is no molecular study that includes a representative from all families of insectivores, and there are no molecular sequence data (data banks or published accounts) for solenodons.

Here, we report a molecular phylogenetic analysis involving all families of insectivores, using complete sequences of the mitochondrial 12S rRNA, 16S rRNA, and tRNA-Valine genes. These data, along with additional sequences from four disparate nuclear

genes, are used to examine the extent of insectivore paraphyly or polyphyly, the conflicting hypotheses regarding the origin of the family Tenrecidae, and the possibility of an African clade of insectivores.

MATERIALS AND METHODS

The 12S rRNA, tRNA-Valine, and 16S rRNA genes were amplified and sequenced [as described elsewhere refs. 8 and 12] from the following taxa: tenrec (*Echinops telfairi*; Insectivora; AF069540), hairy armadillo (*Chaetophractus villosus*; Xenarthra; U61080, AF069534), three-toed sloth (*Bradypus tridactylus*; Xenarthra; AF038022, AF069535), tube-nosed fruit bat (*Nyctimene robinsoni*; Chiroptera; U93061, AF069536), flying fox (*Pteropus hypomelanus*; Chiroptera; U93073, AF069537), false vampire bat (*Megaderma lyra*; Chiroptera; AF069538), mole (*Scalopus aquaticus*; Insectivora; AF069539), solenodon (*Solenodon paradoxus*; Insectivora; AF076646) and capybara (*Hydrochaeris hydrochaeris*; Rodentia; U61081, AF069533). These were combined with sequences already available in GenBank. Accession numbers for many of these sequences are tabulated elsewhere (9); here we list only those accession numbers that did not appear in that recent tabulation. In addition to the above species, the mitochondrial data set included the following taxa: aardvark (*Orycteropus afer*; Tubulidentata), long-nosed armadillo (*Dasypus novemcinctus*; Xenarthra; Y11832), elephant shrew (*Elephantulus rufescens*; Macroscelidea), golden mole (*Amblysomus hottentotus*; Insectivora), African elephant (*Loxodonta africana*; Proboscidea), manatee (*Trichechus manatus*; Sirenia), hyrax (*Procavia capensis*; Hyracoidea), Indian rhino (*Rhinoceros unicornis*; Perissodactyla; X97336), white rhino (*Ceratotherium simum*; Perissodactyla; Y07726), horse (*Equus caballus*; Perissodactyla), harbor seal (*Phoca vitulina*; Carnivora), gray seal (*Halichoerus grypus*; Carnivora), cat (*Felis catus*; Carnivora), fin whale (*Balaenoptera physalus*; Cetacea), blue whale (*Balaenoptera musculus*; Cetacea), cow (*Bos taurus*; Artiodactyla), deer (*Odocoileus virginianus*; Artiodactyla; M35874), mouse deer (*Tragulus napu*, Artiodactyla, M55539), pangolin (*Manis* sp.; Pholidota), shrew (*Sorex palustris*; Insectivora), human (*Homo sapiens*; Primates), gorilla (*Gorilla gorilla*; Primates), orangutan (*Pongo pygmaeus*; Primates), gibbon (*Hylobates lar*; Primates; U99256), rabbit (*Oryctolagus cuniculus*; Lagomorpha; AJ001588), hedgehog (*Erinaceus europaeus*; Insectivora), mouse (*Mus domesticus*; Rodentia), rat (*Rattus norvegicus*; Rodentia), guinea pig (*Cavia porcellus*; Rodentia), monito del monte (*Dromiciops gliroides*; Marsupialia), wallaroo (*Macropus robustus*; Marsupialia; Y10524), bandicoot (*Echymipera kalubu*; Marsupialia), opossum (*Didelphis virginiana*; Marsupialia), and platypus (*Ornithorhynchus anatinus*; Monotremata), for a total of 43 species encompassing 16 placental orders (all eutherians except Dermoptera and Scandentia), 4 marsupial orders, one monotreme, and all insectivore families.

Initial sequence alignments were constructed using CLUSTAL W and then were refined by hand following secondary structure models (13). Regions in which alignments were ambiguous (complex indels) were eliminated from phylogenetic analyses. This resulted in 2,086 nt positions for 43 taxa. All phylogenetic trees involving this data set were rooted at platypus. Upon completion of analyses involving the total mitochondrial data set, these data were analyzed with the African taxa as ingroup, using long-nosed armadillo, cow, rat, and human as outgroups in separate analyses. Again, these alignments were constructed using CLUSTAL W and then refined by hand. Ambiguous regions again were removed; however, a consequence of eliminating 35 taxa from the sequence comparisons was that a larger number of sequence alignment positions could be considered as unambiguous and thus included in the ingroup analyses. A further consequence is a slightly different number of sequence positions represented in alignments involving each of the four outgroups (armadillo: 2,506; cow: 2,624; rat: 2,603; human: 2,560).

To provide a corroborating test of our mitochondrial results, we also obtained sequences for the representative tenrec from two nuclear genes: exon 28 of the gene for von Willebrand factor (vWF; AF076478) and the α-2B adrenergic receptor gene (A2AB; Y17692). New sequence data were also obtained for vWF from mole (AF076479), and hairy armadillo (AF076480). These sequences were combined with already existing data for both of these genes arising from previous studies (8, 9, 14). For vWF, this included the following taxa: human, galago (*Galago crassicaudatus*; Primates), agouti (*Dasyprocta agouti*; Rodentia), rat, mole-rat (*Spalax polinicus*; Rodentia), African elephant, Asian elephant (*Elephas maximus*), dugong (*Dugong dugon*; Sirenia), hyrax, aardvark, elephant shrew, harbor porpoise (*Phocoena phocoena*; Cetacea), cow, pig (*Sus scrofa*; Artiodactyla), hedgehog, golden mole, white rhinoceros, horse (*Equus asinus*), rabbit, pangolin, false-vampire bat, bar-backed fruit bat (*Dobsonia moluccensis*), cat, dog (*Canis familiaris*; Carnivora), flying lemur (*Cynocephalus variegatus*; Dermoptera), tree shrew (*Tupai glis*; Scandentia), and three-toed sloth (scientific name indicated only when it has not been listed already for the mitochondrial data set). This resulted in a 30-taxon data set, representing all 18 eutherian orders and 4 of the 6 families of insectivores. For A2AB, the tenrec was combined with the following taxa: human, guinea pig, rat, mouse, Asian elephant, dugong, hyrax, aardvark, elephant shrew (*Macroscelides proboscideus*), cow, hedgehog, golden mole, mole (*Talpa europaea*), horse, and rabbit (scientific name listed only when that particular species is unique to the A2AB data set). The resulting A2AB alignment included 16 taxa encompassing 11 placental orders and the same 4 families of insectivores. Accession numbers for taxa involving both these genes can be found tabulated elsewhere (9). PCR and sequencing for these two loci were as described elsewhere (8, 14). Phylogenetic trees involving both of these genes were rooted with rodents; alternative rootings produced very similar results.

All of the foregoing sequence alignments are available from the European Molecular Biology Laboratory file server under the following accession numbers: vWF: DS34810; A2AB: DS34808; 12S–16S, 43 taxa: DS34832; 12S–16S, armadillo: DS34831; 12S–16S, cow: DS34811; 12S–16S, rat: DS34812; 12S–16S, human: DS34809.

Using de Queiroz's bootstrap criterion (15) as a measure of data set incongruence, the mitochondrial and the nuclear genes were judged not to be heterogenous (i.e., no conflicting nodes receiving >95% bootstrap support), and thus various data set combinations also were analyzed. These included the following: A2AB + vWF (15 taxa: human, cow, horse, hedgehog, mole, golden mole, tenrec, elephant, hyrax, dugong, aardvark, elephant shrew, rat, caviomorph rodent–guinea pig/agouti, and rabbit), vWF + mtDNA (23 taxa: human, cow, horse, hedgehog, mole, golden mole, tenrec, elephant, hyrax, dugong, aardvark, elephant shrew, rat, guinea pig/agouti, rabbit, false vampire bat, hybrid fruit bat–*Dobsonia*/*Pteropus*, white rhino, hairy armadillo, sloth, cat, pangolin, and hybrid cetacean–porpoise/fin whale), and vWF + A2AB + mtDNA (same 15 taxa as vWF + A2AB). In combined nuclear and mitochondrial data sets, the mitochondrial sequence positions represented were the same as those in the 43-taxon data set. This set of combined data sets allows maximum taxa overlap while also providing alternative perspectives on insectivore ancestries by using different numbers and combinations of taxa. All analyses involving these data set combinations were rooted with rodents.

An examination of the molecular sequence databases reveals that the only other available molecular data for the family Tenrecidae are α and β hemoglobin amino acid sequences from tenrec (*Echinops telfairi*). A variety of other mammal sequences are available for theses two peptides, many of them the same species included in this study. A tandem alignment of both of these peptides was constructed for 14 taxa including the following: human (Swiss Prot accession numbers P01922; P02023), cow (P01966; P02070), horse (P01958; P02062), rabbit (P01948; P02057), African elephant (P01955; P02085), manatee (P07414; P07415), hyrax (P01957; P02086), *Antrozous pallidus* (microbat,

P14387; P14388), rat (P01946; P02091), guinea pig (P01947; P02095), tenrec (P24291; P24292), shrew (P01950; P02060), mole (P01951; P02061), and hedgehog (P01949; P02059). Combined amino acid data sets involving A2AB, vWF, and the two hemoglobins also were constructed. In this case, the alignment consisted of the following 15 taxa: human, rat, guinea pig, elephant, dugong, hyrax, aardvark, golden mole, tenrec, elephant shrew, cow, horse, rabbit, mole, and hedgehog. In this instance, aardvark, golden mole, and elephant shrew were coded as missing data for the hemoglobins because sequences for these taxa are not available in the databases. All trees were rooted with rodents.

Maximum likelihood, minimum evolution, neighbor joining, and parsimony (all data unweighted and transversions only) were used to construct phylogenetic trees. PAUP 4.0 version d54–63 (written by D.L. Swofford) and PHYLIP (15) were used in analysis, with the exception of maximum likelihood analyses of amino acids as well as the vWF + mtDNA data set, which used PUZZLE (16). Bootstrapping was used as a measure of clade strength. All neighbor joining, parsimony, and minimum evolution analyses involved 500 replicates of the data; DNAML (15) maximum likelihood analyses involved 100 replicates. In several instances, statistical tests were conducted to assess the likelihood of constrained topologies relative to the most parsimonious trees. These included winning sites (18), Templeton (19), and Kishino–Hasegawa (20) tests and were implemented by using PAUP. Kishino–Hasegawa tests also were used to evaluate the likelihood of constrained topologies relative to the maximum likelihood trees. Maximum likelihood analyses (DNAML) used empirical base frequencies and a transition:transversion ratio of 2:1, assumed equal rates between sites, and used the global branch swapping option. Neighbor joining trees were estimated using maximum likelihood, logdet and Tamura-Nei–transversion distances. For the parsimony analyses, we used full heuristic searches, with 50 random input orders.

RESULTS

Mitochondrial Genes. All of the different types of phylogenetic analyses involving the mitochondrial data supported a clade of African-origin mammals inclusive of elephant, manatee, hyrax [these three orders generally are termed the Paenungulata; however, see comments in *Discussion* as well as in McKenna and Bell (21) for additional clarification and history of the concept], aardvark, elephant shrew, golden mole, and tenrec (Fig. 1; Table 1). Bootstrap support for this clade ranged from 56% for the very conservative transversion parsimony to 84% for neighbor joining with maximum likelihood distances. The remaining insectivores, hedgehog, solenodon, mole, and shrew were excluded from this clade in all cases. Three of these non-African families, represented by mole, shrew, and solenodon, formed a clade in the majority of tests but generally not with convincing bootstrap support. The hedgehog was in all cases on its own, ambiguously placed in various positions in the tree, depending on method of analysis. Among the non-African insectivores, the mole and shrew invariably were associated together, with moderate bootstrap support. The best tree that supported the monophyly of these non-African families added 12 substitutions to the most parsimonious tree, and this topology was judged not to be significantly different than that MP tree (Templeton: $P = 0.377$; Kishino–Hasegawa: $P = 0.324$; winning sites: $P = 0.419$). Various other nodes in the resulting trees were supported strongly and agreed with now well accepted phylogenetic hypotheses, including for example, a monophyletic Chiroptera, a Paenungulata, and an association of Artiodactyla and Cetacea (see Fig. 1 for example topology).

Within the African clade, all analyses supported a grouping of tenrec and golden mole. In the 43-taxon data set, this support ranged from 39 to 90%. However, with just the Africans as ingroup, when we used four different outgroups, support for this association was generally at our near 100% (Table 1; Fig. 3). The reason for this difference between the ingroup analyses and the

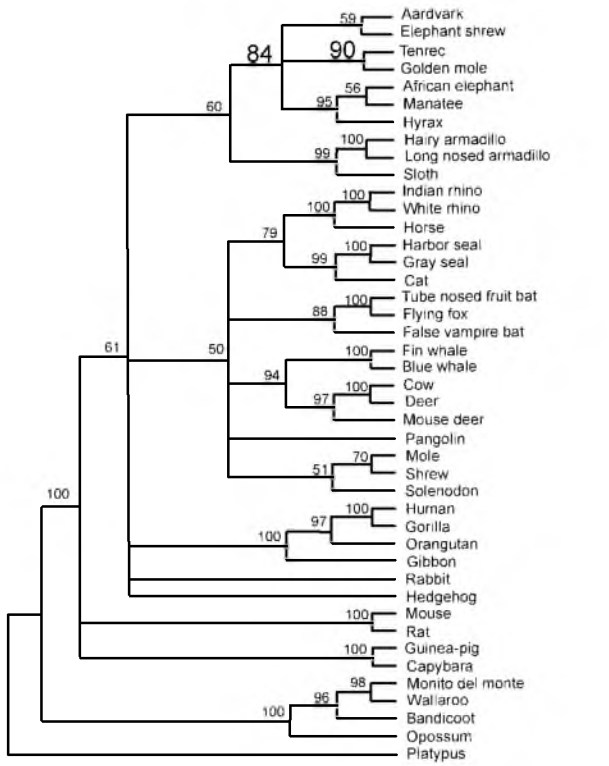


FIG. 1. Majority-rule neighbor joining bootstrap tree based on maximum likelihood distances of the 12S rRNA, tRNA-Valine, and 16S rRNA mitochondrial DNA sequences.

complete data set is that, with the inclusion of only eight taxa, there were fewer ambiguous regions in the alignment, and thus more sequence positions could be included. This, in turn, was a consequence of having only one divergent taxon to align against the Africans and, more importantly, because the Africans clearly exhibited sequence similarity in the most variable regions of the alignment. The similarity of the golden mole and tenrec sequences is further evident in alignments that exclude the outgroup. Although it is uncertain which taxon or taxa represent the most ancient branch of the African clade, the bulk of our analyses suggest that it is the elephant shrew (9). Analyses involving just the Africans, without a non-African outgroup, but instead with trees rooted at elephant shrew, resulted in golden mole/tenrec support of 100% for all analyses. In fact, all possible rootings other than the two African insectivores resulted in 100% support for golden mole/tenrec.

In addition to this strong bootstrap support for the association of golden moles and tenrecs with an all-African clade, statistical tests also support their distinction from the rest of Insectivora. Indeed, all three parsimony tests reject insectivore monophyly at $P < 0.017$ (Table 2). Similarly, the Kishino–Hasegawa test judged insectivore monophyly to be significantly worse than the highest likelihood tree (highest likelihood tree: $-\ln L = 31{,}072.10$; best tree supporting insectivore monophyly: $-\ln L = 31{,}187.73$; $P = 0.0091$). In addition to this ordinal distinction, the morphologically based concepts of the suborder Soricomorpha also were not supported. Butler's (4) concept has mole and shrew forming a clade, joined next by solenodon, and lastly by golden mole/tenrec. This topology costs an additional 38 substitutions and was judged significantly worse than the MP tree using all three tests (Table 2; concept of Soricomorpha was not evaluated statistically using Kishino–Hasegawa maximum likelihood test). MacPhee and Novacek (3), on the other hand, suggest Soricomorpha is comprised of tenrecs, solenodons, shrews, and moles, relative relationships undefined. This arrangement, distinct from Butler's in that golden moles are excluded, costs an additional 52 substitu-

Table 1. Bootstrap support for Afrotheria and Afrosoricida monophyly

	Level of support for monophyly of Afrotheria	Level of support for monophyly of Afrosoricida
Mitochondrial DNA (43 taxa)		
Parsimony	74	39 (98, 96, 75, 95)
Transversion parsimony	56	79 (92, 96, 55, 100)
Neighbor joining		
ML	84	90 (99, 98, 99, 100)
Logdet	82	87 (99, 97, 97, 100)
Tamura-Nei (TV)	69	75 (86, 96, 90, 94)
Maximum likelihood	77*	68* (97, 94, 98, 99)
A2AB (16 taxa)		
Parsimony	100	27
Transversion parsimony	99	36
Neighbor joining		
ML	100	25
Logdet	100	23
Tamura-Nei (TV)	93	29
Maximum likelihood	100	43
vWF (30 taxa)		
Parsimony	78	45
Transversion parsimony	80	58
Neighbor joining		
ML	77	6
Logdet	82	7
Tamura-Nei (TV)	67	26
Maximum likelihood	81†	30†
A2AB + vWF (15 taxa)		
Parsimony	100	55
Transversion parsimony	99	57
Neighbor joining		
ML	100	27
Logdet	100	28
Tamura-Nei (TV)	100	35
Maximum likelihood	100	60
Neighbor joining, amino acids	100	84
Amino acid parsimony	100	61
Amino acid maximum likelihood	96	74
A2AB + mtDNA (15 taxa)		
Parsimony	100	65
Transversion parsimony	100	68
Neighbor joining		
ML	100	89
Logdet	100	83
Tamura-Nei (TV)	100	82
Maximum likelihood	100	89
vWF + mtDNA (23 taxa)		
Parsimony	99	70
Transversion parsimony	99	94
Neighbor joining		
ML	100	63
Logdet	99	69
Tamura-Nei (TV)	99	86
Maximum likelihood	89	83
mtDNA + vWF + A2AB (15 taxa)		
Parsimony	100	76
Transversion parsimony	100	90
Neighbor joining		
ML	100	65
Logdet	100	71
Tamura-Nei (TV)	100	82
Minimum evolution – ML	100	83
Maximum likelihood	100	84
A2AB + vWF + α-β hemoglobin (15 taxa)		
Parsimony	98	87
Neighbor joining	99	88
Maximum likelihood	97	72

Numbers in parentheses indicate the results of analyses with Africans as ingroup, using armadillo, cow, rat, and human, respectively, as outgroups. Number of taxa were reduced to 24 for analysis indicated with *; this set of 24 included all of the insectivores, all of Afrotheria, in addition to the following: mouse, rat, rabbit, human, pangolin, cat, gray seal, horse, cow, fin whale, false vampire bat, long nosed armadillo, and opossum. Number of taxa was reduced to 21 for analysis indicated with † including human, rabbit, elephant shrew, golden mole, hedgehog, tenrec, aardvark, dugong, hyrax, Asian elephant, sloth, mole, tree shrew, pangolin, cow, porpoise, horse, dog, *Dobsonia*, flying lemur, and rat.

tions and also was judged significantly different than the MP tree (Table 2).

Nuclear Genes and Combined Data Sets. Both the vWF and A2AB genes were congruent with the mitochondrial data in supporting an African clade inclusive of aardvark, elephant shrew, the paenungulates, golden mole and tenrec, at the exclusion of the other insectivores in these data sets (mole and hedgehog). Bootstrap support for this African association ranges from 93 to 100% for A2AB and 67 to 82% for vWF. Like the mitochondrial data, parsimony and maximum likelihood statistical tests rejected insectivore monophyly (Table 2; A2AB-highest likelihood tree: $-\ln L = 7{,}682.90$; best tree supporting insectivore monophyly: $-\ln L = 7{,}800.74$; $P < 0.0001$; vWF-highest likelihood tree: $-\ln L = 15{,}996.64$; best tree supporting insectivore monophyly: $-\ln L = 16{,}122.15$; $P < 0.0001$). Unlike the mtDNA data set, which did not result in a clear phylogenetic placement for the hedgehog, the nuclear genes positioned it along with the mole, the other non-African insectivore in these data sets. Joining the African insectivores with the mole, in other words the construction of a partial Soricomorpha (i.e., solenodon and shrew sequences not yet available for the nuclear genes), costs an additional 34/38 and 29/54 substitutions for A2AB and vWF, respectively, all such constrained topologies judged significantly worse than the MP trees (Table 2).

Combining these two nuclear genes in a concatenated 15-taxon alignment increased African bootstrap support to 100% (or very nearly) for all analyses and resulted in some support for a golden mole/tenrec clade (Table 1). This latter grouping exhibited minimal (nucleotide sequence level) to moderate (amino acid sequence level) bootstrap support (Table 1). The classical concepts of a monophyletic Insectivora or Soricomorpha were highly significantly rejected (Table 2; highest likelihood tree: $-\ln L = 17{,}075.46$; best tree supporting insectivore monophyly: $-\ln L = 17{,}264.93$; $P < 0.0001$). Combining nuclear with mitochondrial data yields 100% bootstrap support for the African superordinal clade, at the exclusion of the non-African insectivores, which form a distinct well supported clade (Fig. 2) and with the majority of tests supporting an association of golden mole and tenrec at 80–95% bootstrap. Parsimony statistical tests rejected both insectivore and soricomorph monophyly at $P < 0.0001$ (Table 2), and maximum likelihood tests similarly rejected the concept of Insectivora at $P < 0.0001$ ($-\ln L$ values for highest likelihood tree and best tree supporting insectivore monophyly, respectively: A2AB + mtDNA, 22,074.83/22,316.80; vWF + mtDNA, 33,562.89/33,823.98 and vWF + A2AB + mtDNA, 31,831.15/32,152.60).

Hemoglobin analyses also supported a distinct evolutionary history for the tenrecs and the rest of Insectivora. Amino acid sequence data for these peptides include four families of Insectivora: tenrec, shrew, mole, and hedgehog. The most parsimonious tree and the NJ tree both supported a clade of shrew, mole, and hedgehog. In all cases, the tenrec was separated from that group, joining instead one (in the case of parsimony) or all of the Paenungulates. Bootstrap support for these associations was generally low; however, combining these hemoglobin data with amino acid sequences from A2AB and vWF resulted in strong support for the monophyly of the African clade (Table 2), a golden mole/tenrec grouping inside the African clade (Table 2), and the common ancestry of mole and hedgehog (the two other insectivores in this data set combination; 100% for each of parsimony, NJ and maximum likelihood). Statistical tests for this data set combination rejected insectivore monophyly, as well as the concept of Soricomorpha (Table 2).

Table 2. Results of statistical tests regarding Insectivora and Soricomorpha monophyly

Phylogenetic loci	Score of the MP tree (RI)	Parsimony score of alternative hypotheses (RI)			Statistical tests of alternative hypotheses								
					Isectivora			Soricomorpha					
		Insectivora	Soricomorpha					Butler			MacPhee–Novacek		
			B	MN	KH	Temp.	WS	KH	Temp.	WS	KH	Temp.	WS
mtDNA	5729 (0.438)	5769 (0.432)	5767 (0.432)	5781 (0.430)	0.0168	0.0154	0.0168	0.0145	0.0252	0.0119	0.0020	0.0252	0.0119
A2AB	1274 (0.474)	1310 (0.433)	1312 (0.430)	1308 (0.435)	0.0006	0.0021	0.0006	0.0002	0.0010	0.0006	0.0002	0.0010	0.0004
vWF	3094 (0.387)	3130 (0.374)	3148 (0.368)	3123 (0.377)	0.0029	0.0068	0.0084	<0.0001	0.0001	<0.0001	0.0261	0.0507	0.0218
A2AB + vWF	3020 (0.357)	3085 (0.322)	3100 (0.314)	3078 (0.326)	0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
A2AB + mtDNA	3766 (0.309)	3828 (0.283)	3836 (0.280)	3874 (0.264)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
vWF + mtDNA	6225 (0.310)	6288 (0.297)	6314 (0.292)	6299 (0.295)	<0.0001	<0.0001	0.0006	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
mtDNA + vWF + A2AB	5560 (0.315)	5652 (0.289)	5675 (0.282)	5663 (0.286)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
A2AB + vWF + α-β hemoglobin	1540 (0.376)	1553 (0.358)	1578 (0.324)	1573 (0.331)	0.0325	0.0624	0.0485	<0.0001	<0.0001	<0.0001	0.0002	0.0011	0.0003

B, Butler's concept of Soricomorpha; MN, MacPhee and Novacek's concept of Soricomorpha; RI, retention index.

DISCUSSION

The results reported here, from several disparate molecular loci, congruently support the view that tenrecs, in addition to golden moles (8), are members of the "African clade" of mammalian orders and do not belong within the classical order Insectivora. Golden moles and tenrecs are therefore more closely related to aardvarks and elephants than they are to shrews, moles, and hedgehogs. Because this African clade is well supported both from a statistical and congruence perspective but is not currently recognized taxonomically, we propose the superordinal name Afrotheria. Furthermore, the golden moles and tenrecs form an evolutionary clade within Afrotheria, and thus it is necessary to propose a new order to accommodate those taxa. For this association of African insectivore families, we propose the name "Afrosoricida" (African shrew-like mammals). Two extinct orders of mammals, Desmostylia and Embrithopoda, are believed to be closely related to the paenungulate orders (21, 22) and therefore should be included in Afrotheria based on available phylogenetic evidence. Even so, these extinct taxa may not have originated in Africa but rather from ancestors that radiated out of Africa. The oldest embrithopods are the Paleocene genera *Phenacolophus* and *Minchenella* from Asia whereas the oldest African embrithopods are Eocene (21). The oldest desmostylian fossils are from the Oligocene, and the entire fossil record for this group (six genera) is confined to the northern Pacific Rim (21). The place and time of origin of aardvarks are also uncertain; they are presumed to be of African origin, but there are no unequivocal Paleogene fossils referable to this group.

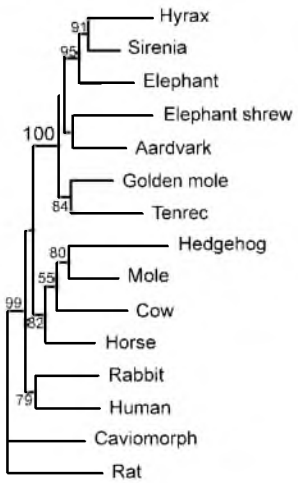


FIG. 2. Maximum likelihood tree of the combined 12S rRNA, tRNA-Valine, 16S rRNA, vWF, and A2AB data set, with branches drawn proportional to amount of sequence change, and with maximum likelihood bootstrap values added; only those bootstrap figures in excess of 50% are indicated.

Anatomical studies have not yet provided support for such a diverse African clade, and indeed there is not a single morphological synapomorphy that defines Afrotheria. There are, however, morphological characters that support a possible common ancestry of golden moles and tenrecs, although not to the exclusion of the other insectivores. These include zalambdodont cheekteeth (also in *Solenodon*), a basisphenoid contribution to the bulla (also in Erinaceidae), and orthomesometrial implantation of the blastocyst (3). There are also data from albumin immunodiffusion comparisons that join golden moles and tenrecs in a distinct clade, separate from all other insectivores (23). These other two sources of data are therefore at least partially congruent with our proposal for an Afrosoricida. By using the recently established linear relationships describing change in 12S rRNA Tamura Nei relative rate-adjusted transversion distances with time (24), we estimate that the ancestor to this new order resided in Africa somewhere $\approx$70 million years ago (mya) and that the split between tenrecs and golden moles occurred between 50 and 57 million years age. We suggest, therefore, that this proposed new order, similar to other recent estimates regarding other placental orders, may predate the Cretaceous–Tertiary boundary (8, 24–27).

In addition to the lack of support for a monophyletic Insectivora, the subordinal concepts of Butler (4) and MacPhee and Novacek (3) are also not supported by our data because of the different ancestry for the golden moles and tenrecs. Although MacPhee and Novacek's hypothesis envisions a possibly very different pattern of evolutionary history for the golden moles, and as such they distinguish them from Soricomorpha, they find no such reason to exclude tenrecs. Our data, on the other hand, reject not only golden moles but also tenrecs as having any

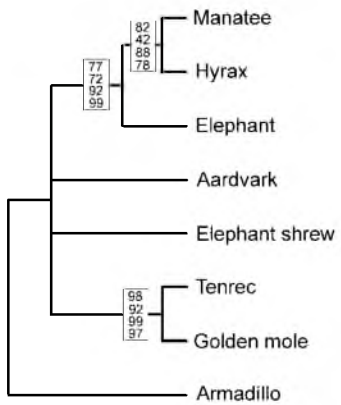


FIG. 3. Strict consensus bootstrap tree of parsimony, transversion parsimony, neighbor joining (with ML distances), and maximum likelihood analyses (bootstrap figures listed from top to bottom respectively) of the 12S rRNA, tRNA-Valine, and 16S rRNA data set, rooted with armadillo.

association with the remaining soricomorph groups, solenodon, mole, and shrew. The morphological concept of Soricomorpha in fact is based largely on the considerable differences between that suborder and hedgehogs (Erinaceomorpha). Our results, on the other hand, suggest that hedgehogs are much more closely related to other soricomorphs such as moles than are tenrecs or golden moles. Although the mitochondrial data alone do not group hedgehogs with the other non-African insectivores, the other three families do form a monophyletic group. All other data sets and combinations of data sets (including those involving mtDNA) associate hedgehog with the non-African insectivore, mole, generally with high bootstrap support (see Fig. 2 for example). These data, taken collectively, suggest that the hedgehogs, shrews, moles, and solenodons form a monophyletic group and as such should be retained in the order Insectivora.

Our demonstration of an evolutionary association of golden moles and tenrecs within an African superordinal clade has important ramifications for the evaluation of character state evolution in these and other mammals. It means, for example, that zalambdodonty evolved independently in Afrosoricida and solenodons among living insectivores. Similarly, other features currently used to define the monophyly of Insectivora, such as (*i*) hindgut simplification, with the correlated absence of cecum, (*ii*) reduction of the pubic symphysis, and (*iii*) large maxillary contribution to the orbit, probably represent convergences rather than shared-derived traits. The converse alternative [that there is molecular convergence between mitochondrial rRNA genes, as well as a number of nuclear, protein-encoding genes, of highly disparate function located on different chromosomes, including a blood clotting factor (van Willebrand factor), another involved in the neural circuitry regulating cardiovascular function (A2AB) and hemoglobins] is unparsimonious in the extreme.

These results highlight the important role of plate tectonics and biogeography in the early diversification of placental mammals. Nearly one-third of the extant orders of placental mammals belong to the superorder Afrotheria, and molecular clock estimates suggest an origin for that lineage in the mid-Cretaceous (105–90 mya) at a time when the continent was isolated (8, 24–26). When representatives of most of the extant placental orders first appear in the fossil record in the early Cenozoic (65–55 mya), many of the continents had at least periodic connection, and thus the distributions of those fossils may not accurately reflect the land areas where those groups arose (25). The challenge for paleontologists will be to locate more Cretaceous fossils of placentals to better understand the historical biogeography of mammals.

We thank Morris Goodman and Jose Ottenwalder for providing some samples, David Swofford for permission to use PAUP 4.0D54-63, and Alan Walker for suggestion of the superordinal name. This work was supported by the following grants: the Nuffield Foundation and the Royal Society to M.J.S., a European Commission -TMR grant (ERB-FMRX-CT98-0221) to M.J.S. and W. W. de Jong; and by the National Science Foundation (DEB-9419617), an Alfred P. Sloan Young Investigator's Award in Molecular Evolution, and an intramural grant from the Academic Senate at the University of California, Riverside to M.S.S.

1. Huxley, T. H. (1880) *Proc. R. Soc. Lond.* **43,** 649–662.
2. Matthew, W. D. (1909) *Mem. Am. Mus. Nat. Hist.* **9,** 291–567.
3. MacPhee, R. D. E. & Novacek, M. J. (1993) in *Mammal Phylogeny*: *Placentals*, eds. Szalay, F. S., Novacek, M. J. & McKenna, M. C. (Springer, New York), Vol. 2, pp. 13–31.
4. Butler, P. M. (1988) in *The Phylogeny and Classification of the Tetrapods*: *Mammals*, ed. Benton, M. J. (Clarendon, Oxford), Vol. 2, pp. 117–141.
5. Broom, R. (1916) *Proc. Zool. Soc. London* **1916,** 449–459.
6. Heim de Balsac, H. & Bourlière F. (1954) in *Traite de Zoologie*, ed. Grassé, P. P. (Masson & Cie, Paris), Vol. 17, pp. 1653–1679.
7. Butler, P. M. (1972) in *Studies in Vertebrate Evolution*, eds. Joysey, K. A. & Kemp, T. S. (Oliver and Boyd, Edinburgh), pp. 253–265.
8. Springer, M. S., Cleven, G. C, Madsen, O., de Jong, W. W., Waddell, V. G., Amrine, H. M. & Stanhope, M. J. (1997) *Nature (London)* **388,** 61–64.
9. Stanhope, M. J., Madsen, O., Waddell, V. G., Cleven, G. C., de Jong, W. W. & Springer, M. S. (1998) *Mol. Phylogenet. Evol.*, in press.
10. Carroll, R. L. (1988) *Vertebrate Paleontology and Evolution* (Freeman, New York).
11. Benton, M. J., ed. (1993) *The Fossil Record* (Chapman & Hall, New York), Vol. 2.
12. Springer, M. S., Hollar, L. J. & Burk, A. (1995) *Mol. Biol. Evol.* **12,** 1138–1150.
13. Springer, M. S. & Douzery, D. (1996) *J. Mol. Evol.* **43,** 357–373.
14. Porter C. A., Goodman, M., & Stanhope, M. J. (1996) *Mol. Phylogent. Evol.* **5,** 89–101.
15. de Queiroz, A. (1993) *Syst. Biol.* **42,** 368–372.
16. Felsenstein, J. (1993) PHYLIP (phylogeny inference package) (Department of Genetics, University of Washington, Seattle), Ver. 3.5c.
17. Strimmer, K. & von Haeseler, A. (1996) *Mol. Biol. Evol.* **13,** 964–969.
18. Prager, E. M. & Wilson, A. C. (1988) *J. Mol. Evol.* **27,** 326–335.
19. Templeton, A. R. (1983) *Evolution* **37,** 221–224.
20. Kishino, H. & Hasegawa, M. (1989) *J. Mol. Evol.* **29,** 170–179.
21. McKenna, M. C. and Bell, S. K. (1997) *Classification of Mammals Above the Species Level* (Columbia Univ. Press, New York).
22. Benton, M. J. (1997) *Vertebrate Palaentology.* (Chapman & Hall, London).
23. Sarich, V. M. (1993) *Mammal Phylogeny*: *Placentals*, eds. Szalay, F. S., Novacek, M. J. & McKenna, M. C., pp. 103–114 (Springer, New York), Vol. 2.
24. Springer, M. S. (1998) *J. Mamm. Evol.* **4,** 285–302.
25. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. (1996) *Nature (London)* **381,** 226–229.
26. Kumar, S. & Hedges, S. B. (1998) *Nature (London)* **392,** 917–920.
27. Janke, A., Xu, X. & Arnason, U. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 1276–1281.

Chapter Six

Parallel adaptive radiations in two major clades of placental mammals

Ole Madsen, Mark Scally, Christophe J. Douady, Diana J. Kao, Ronald W. DeBry, Ronald Adkins, Heather M. Amrine, Michael J. Stanhope, Wilfried W. de Jong and Mark Springer. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* **409**: 610-614

Higher level relationships among placental mammals, as well as the historical biogeography and morphological diversification of this group, remain unclear[1–3]. Here we analyse independent molecular data sets, having aligned lengths of DNA of 5,708 and 2,947 base pairs, respectively, for all orders of placental mammals. Phylogenetic analyses resolve placental orders into four groups: Xenarthra, Afrotheria, Laurasiatheria, and Euarchonta plus Glires. The first three groups are consistently monophyletic with different methods of analysis. Euarchonta plus Glires is monophyletic or paraphyletic depending on the phylogenetic method. A unique nine-base-pair deletion in exon 11 of the *BRCA1* gene provides additional support for the monophyly of Afrotheria, which includes proboscideans, sirenians, hyracoids, tubulidentates, macroscelideans, chrysochlorids and tenrecids. Laurasiatheria contains cetartiodactyls, perissodactyls, carnivores, pangolins, bats and eulipotyphlan insectivores. Parallel adaptive radiations have occurred within Laurasiatheria and Afrotheria. In each group, there are aquatic, ungulate and insectivore-like forms.

The combination of fossil and anatomical data has suggested moderately well-resolved phylogenetic trees for the 18 extant orders of placental mammals[1,2]. DNA sequences have provided consistent support for only a few of the proposed superordinal groups, notably for Paenungulata (elephants, sea cows and hyraxes) and Cetartiodactyla (artiodactyls and whales)[3], and have rejected some traditional clades such as Archonta (primates, tree shrews, flying lemurs and bats)[4]. Molecular data have also suggested new sets of relationships; Afrotheria, a group that includes paenungulates, aardvarks, elephant shrews, golden moles and tenrecs, is supported by both mitochondrial ribosomal RNA and nuclear protein-coding genes[5,6]. These same sequences suggest that lipotyphlan insectivores are paraphyletic or polyphyletic.

Increased resolving power may result from concatenations of individual genes[7]. We concatenated DNA sequences for mitochondrial RNA genes and three nuclear genes (A2AB, IRBP, vWF), including 16 new sequences (see Methods). This data set includes 26 placental taxa, representative of all eutherian orders, and two marsupial outgroups (see Fig. 1A). In addition, golden mole and tenrec, belonging to the insectivore families Chrysochloridae and Tenrecidae, respectively, are represented by sequences for all genes

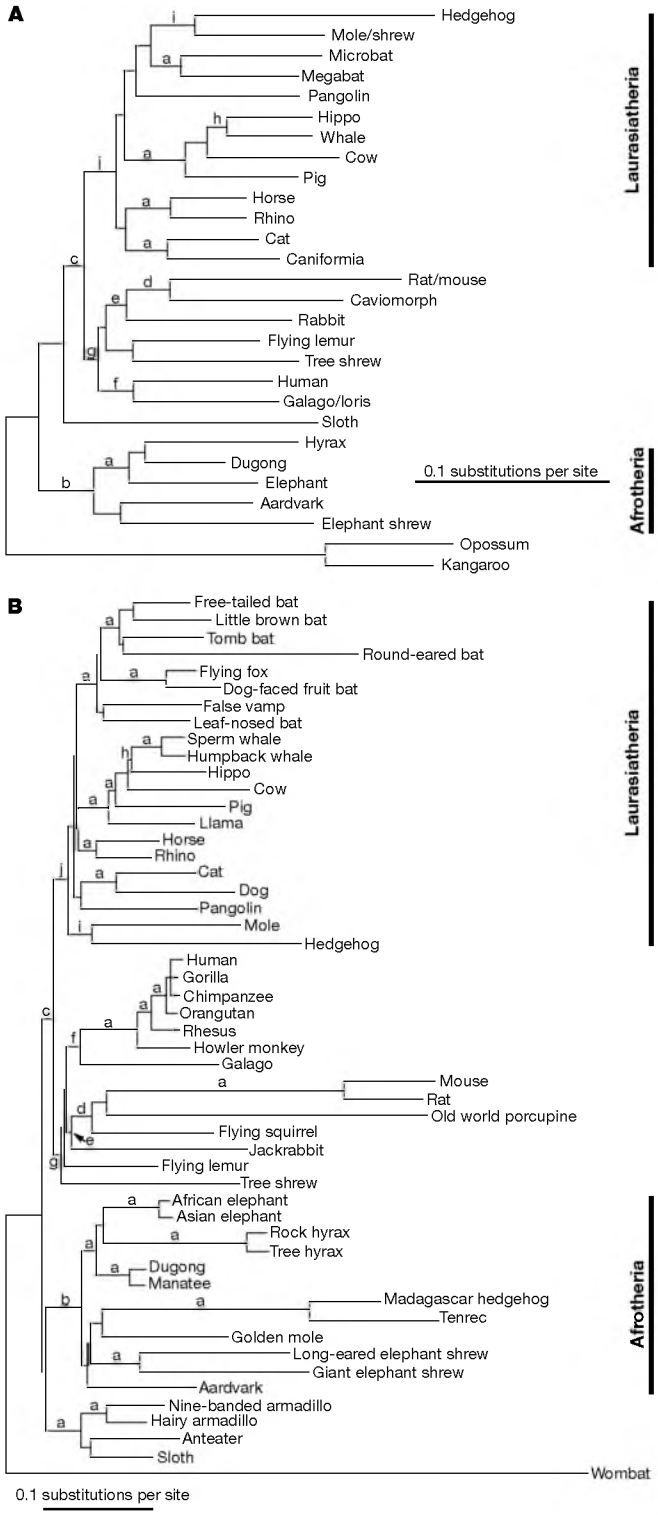


Figure 1 Rooted maximum-likelihood trees. **A**, The 5,708-bp data set. **B**, The 2,947-bp data set. For chimaeric sequences in **A**, the combined species or the higher taxonomic unit are indicated. Branch lengths are proportional to the amount of sequence change (scale bar, 10% sequence divergence). Bootstrap support values for nodes b–j with different phylogenetic methods are shown in Table 1; nodes labelled a are supported at or above 90% by all methods. Afrotheria also includes golden moles and tenrecs, which are not represented by IRBP sequences, and therefore not shown in **A**; in analyses with the 5,708-bp data set that included golden mole and tenrec, bootstrap support for Afrotheria ranged from 88% to 100%. For the 5,708-bp data set, the 12 statistically acceptable branches for the location of the root, in decreasing order of likelihood score, are branch b (as shown), sloth, rabbit, branch leading to rat/mouse, branch c, branch d, branch e, caviomorph, branch j, tree shrew, branch leading to flying lemur + tree shrew, and galago/lorls; other branches were rejected. For the 2,947-bp data set, the statistically acceptable options, in decreasing order of likelihood score, were branch c (as shown), branch b, and the base of Xenarthra. The name Laurasiatheria (''from the area of Laurasia or Europe + Asia + North America'')[16] was suggested for this clade, but no support in the form of bootstrap analyses and/or statistical tests was provided.

except IRBP; inclusion of these taxa in a subset of analyses was necessary to investigate the proposed diphyly of the insectivores and to confirm the naturalness of Afrotheria. In all, 5,708 aligned nucleotide positions were used in phylogenetic analyses.

Hypothesis testing with independent data sets is fundamental in phylogenetics. To establish a data set that was independent of the 5,708-bp concatenation, we sequenced roughly 3 kilobases (kb) of exon 11 of the single-copy breast and ovarian cancer susceptibility gene 1 (*BRCA1*) for 33 taxa and combined these with 19 GenBank sequences (see Methods). The aligned BRCA1 data set includes representatives of all placental orders and one marsupial outgroup (for names, see Fig. 1B). Finally, we combined the 5,708-bp and 2,947-bp data sets (total 8,655 bp) for 26 overlapping placentals and one marsupial.

Phylogenetic trees (maximum likelihood) for the 5,708-, 2,947- and 8,655-bp data sets are shown in Fig. 1A and B, and Supplementary Information, respectively. Bootstrap support values for different phylogenetic methods are given in Table 1 for the 5,708- and 2,947-bp data sets, and in Supplementary Information for the 8,655-bp data set. In spite of the differences in taxonomic sampling, as well as some conspicuously long branches on the BRCA1 tree, placental orders are resolved into the following four groups on all three maximum-likelihood trees: Xenarthra, Afrotheria, Laurasiatheria, and Glires (rodents and lagomorphs) + Euarchonta (flying lemurs, tree shrews and primates). Xenarthra, Afrotheria, and Laurasiatheria are robust with different phylogenetic methods (Fig. 1B; Table 1; Supplementary Information). Afrotheria is also supported by a 9-bp deletion that is shared by all 12 afrotherians that were included in our study (Fig. 2). The absence of this deletion in other placentals, as well as in marsupials, indicates that this feature may be a shared derived character of Afrotheria. Despite the absence of any morphological evidence for Afrotheria[2,8], it now becomes increasingly difficult to explain this hypothesis except in the context of shared common ancestry. Within Afrotheria, there is strong support for Paenungulata. Our analyses provide the first robust support for the Laurasiatheria clade. Within this clade, there is strong support for Cetartiodactyla and the monophyly of eulipotyphlan insectivores (hedgehogs, shrews and moles).

Support for Euarchonta + Glires was inconsistent, largely because the position of the root was sensitive to the phylogenetic method that was used (Table 1; Supplementary Information). Maximum-likelihood analyses with the BRCA1 data set provided 100% bootstrap support for the monophyly of Glires + Euarchonta, whereas parsimony provided only 18% support. The shortest parsimony trees (two at 10,867 steps; Supplementary Information) for BRCA1 root on Old World porcupine, rendering Euarchonta + Glires, Glires and Rodentia paraphyletic; 35 additional steps are required to root in the same location as the likelihood tree shown in Fig. 1B. Among the individual bootstrap replicate trees for parsimony, rootings occurred on long, topologically separated branches as follows: Old World porcupine (69%); rat-mouse (11%); Madagascar hedgehog-tenrec (11%); tree shrew (4%); hedgehog (3%); and round-eared bat (2%). Maximum likelihood did not root on these long branches. Instead, there was 100% bootstrap support for the monophyly of Laurasiatheria + Euarchonta + Glires (branch c), and individual bootstrap replicate trees, with or without an allowance for rate-heterogeneity, rooted on branch b (base of Afrotheria), branch c or at the base of Xenarthra; these three branches are topologically adjacent, which suggests a strong affinity of the root for this general region of the tree. Furthermore, statistical tests with likelihood reject rooting the BRCA1 tree on all branches except for these three. Finally, Monte Carlo simulations mirrored our results for the actual BRCA1 data set and showed that parsimony will root on long branches and fail to recover the correct root given the topology depicted in Fig. 1B (Supplementary Information).

Twelve roots are statistically acceptable on the likelihood tree for the 5,708-bp data set (Fig. 1A). The three possible BRCA1 roots coincide with three of the five best roots on the 5,708-bp data set tree. These are also the three best roots for the 8,655-bp data set. Although several roots in the Euarchonta + Glires group could not be rejected with the 5,708- and 8655-bp data sets, we do not favour these roots. First, there are numerous morphological synapomorpies for both rodent monophyly[9] and Glires[2]. Second, other molecular results also favour rodent monophyly and even the monophyly of Glires, when phylogenetic methods are used that address problems such as nonstationarity and long-branch attraction[10,11].

Afrotheria and Xenarthra have observed first occurrences in the fossil record on Southern Hemisphere landmasses—Africa in the case of Afrotheria[5,6] and South America in the case of Xenarthra[12]. First occurrences in the fossil record do not permit unequivocal assignment of continental origin. Nevertheless, there is a strong possibility that Afrotheria and Xenarthra have Gondwanan origins. Of the three roots (base of Afrotheria, branch b; base of Laurasiatheria + Euarchonta + Glires, branch c; base of Xenarthra) that are statistically acceptable on all three trees, rooting on the xenarthran branch is the only hypothesis that is consistent with the morphology-based Epitheria hypothesis[2]. At the same time, rooting on the Xenarthra branch results in paraphyly of a possible Gondwanan group (i.e., Xenarthra + Afrotheria) at the base of the placental tree. Rooting on branch b also results in paraphyly of this Southern Hemisphere group. The suggestion that a Gondwanan assemblage is paraphyletic at the base of the tree contrasts with traditional views on the evolution of placental mammals, which place their most recent common ancestor in the Northern Hemisphere[13]. Considering the debate about the timing of the principal placental divergences[14,15], the paucity of Cretaceous mammal fossils from the Southern Hemisphere[15], and the discoveries of tribosphenic fossils in the Southern Hemisphere, including a possible placental[16], a Gondwanan origin for extant placentals should not be excluded. A Southern Hemisphere ancestry for

Table 1 Bootstrap support percentages for branches b–j in Fig. 1 with different phylogenetic methods

	Parsimony				Distance methods						Maximum likelihood			
Node	Unweighted		Transversion		NJ-WAVE		ML-GTR		Logdet		No rate heterogeneity		Rate heterogeneity	
	5,708	2,947	5,708	2,947	5,708	2,947	5,708	2,947	5,708	2,947	5,708	2,947	5,708	2,947
b	100	89	100	83	100	96	100	76	100	94	100	100	100	100
c	53	11	62	29	51	56	19	40	30	39	45	100	64	100
d	79	22	74	30	41	69	99	48	100	61	80	99	94	99
e	53	3	4	26	5	34	42	17	54	15	41	67	59	68
f	97	88	71	83	98	89	99	84	99	85	99	99	96	98
g	48	18	58	30	6	49	5	35	10	34	44	100	66	100
h	93	13	72	42	100	85	100	88	100	88	98	36	96	44
i	95	95	94	100	100	100	76	99	74	99	100	100	96	100
j	96	93	87	97	76	97	81	89	78	98	99	100	99	100

In the first row, 5,708 and 2,947 refer to the 5,708-bp and 2974-bp data sets, respectively. In maximum-likelihood analyses that did not allow for rate heterogeneity, transition/transversion ratios were set ot 1.81 and 2.22 for the 5,708- and 2947-bp data sets, respectively. In maximum-likelihood analyses that allowed for rate heterogeneity, transition/transversion ratios were set ot 1.89 and 2.24 for the 5,708- and 2947-bp data sets, respectively. ML-GTR, maximum likelihood with general time reversible model; WAVE, weighted average distances (see Methods).

crown group placentals also makes the Garden of Eden hypothesis[15] more viable as an explanation for the discrepancies between molecular and paleontological datings for the interordinal diversification of placental mammals[14]. Indeed, in agreement with other molecular studies that place basal placental divergences well before the Cretaceous/Tertiary boundary[11,14], we estimate that the clades Afrotheria and Laurasiatheria diverged from each other during the Early Cretaceous, roughly 111–118 million years ago (Supplementary Information).

Mitochondrial protein coding genes have been used to examine interordinal relationships. A study[17] that included 11 placental orders provides support for a clade that resembles Laurasiatheria and contains cetartiodactyls, perissodactyls, carnivores, bats and mole (pangolin was not represented). The key difference is that hedgehog is dissociated from the other eulipotyphlan (mole), and even from other laurasiatherians, in the mitochondrial protein trees. Thus, our data minimize the sundering of Lipotyphla into two orders, Afrosoricida (golden moles, tenrecs)[6] and Eulipotyphla[18], whereas mitochondrial protein genes additionally contradict eulipotyphlan monophyly[17].

Our molecular results support morphology-based hypotheses such as Paenungulata. Other hypotheses, such as Ungulata (cetartiodactyls, paenungulates, perissodactyls and tubulidentates), Altungulata (paenungulates and perissodactyls), Anagalida (Glires and elephant shrews), Lipotyphla, Archonta, Volitantia (bats and flying lemurs), Artiodactyla and Edentata (xenarthrans and pholidotes) are rejected with statistical tests (Supplementary Information) in favour of Afrotheria and Laurasiatheria. If hypotheses such as Afrotheria and Laurasiatheria are correct, then morphology has failed to recover some of the most fundamental clades in all of Eutheria and has not revealed even a single synapomorphy for these groups. Morphological synapomorphies, if they were present, may have been eroded or overprinted during the subsequent evolutionary history of these groups.

Placental and marsupial mammals underwent parallel adaptive radiations that resulted in spectacular examples of convergence. The current evidence also suggests parallel adaptive radiations in the placental clades Afrotheria and Laurasiatheria. Ungulate-like forms include perissodactyls and artiodactyls in Laurasiatheria and hyracoids in Afrotheria. Extant hyracoids represent only a fraction of the diversity once present in Hyracoidea and extinct forms show convergence in body size and dental morphology towards suids, tapirs, and equoids[19]. Indeed, hyracoids were Africa's dominant small- to medium-sized terrestrial herbivores during the Oligocene[19]. Pangolins (Laurasiatheria) and aardvarks (Afrotheria) both show specializations for ant-eating. Fully aquatic forms include cetaceans in Laurasiatheria and sirenians in Afrotheria. Examples among the insectivores include true moles (Talpidae) in Laurasiatheria versus golden moles (Chrysochloridae) in Afrotheria, terrestrial shrews (Soricidae) in Laurasiatheria versus shrewlike tenrecs (*Microgale*) in Afrotheria, and hedgehogs (Erinaceidae) in Laurasiatheria versus Madagascar 'hedgehogs' (*Echinops*) in Afrotheria. Beyond these comparisons, otters (Carnivora) in Laurasiatheria and African otter shrews (*Potamogale*) in Afrotheria resemble each other in general appearance and semi-aquatic habit. A noteable difference is the absence of flying mammals in Afrotheria.

Matthew[20] suggested that "insectivores" were the central stock from which other eutherian orders originated, and there is abundant evidence for insectivore-like mammals in the fossil record[21]. Even today, living lipotyphlan insectivores and fossil insectivores such as the Cretaceous palaeoryctids are often regarded as among the more primitive eutherian mammals[1,2]. Easteal[22] suggested that primitive placentals from the Cretaceous may have diversified phylogenetically before they diverged morphologically and acquired the diagnostic features of ordinal level crown-group clades. The present results, which deploy living lipotyphlan insectivores into Laurasiatheria and Afrotheria, respectively, are consistent with the view that insectivore-like forms were the central stock from which other groups originated and further suggest that different insectivore-like lineages have independently given rise to diverse eutherian taxa in different geographic venues. The finding that parallel adaptive radiations have occurred within Eutheria also helps to explain the difficulties that have accompanied efforts to elucidate interordinal relationships based on morphology. □

	333333333333333333333333333333
	555555566666666667777777777888
Taxon	345678901234567890123456789O123
Kangaroo	AGTAATGATATTTTAACTTCTGATAACTCCT
Wombat	AGTAGTGACATTTTAGCCTCTGATAACTCCA
Opossum	AGTAATGACATTTTAGCCTCTGATTACTCCT
Aardvark	AGTGATGGCCTG---------GATGGCTCAC
Giant elephant shrew	AGTGATGGCCTC---------GATGATGCCC
Tenrec	AGCCACGGCCTG---------GGTGACTCTC
Golden mole	AGTGATGGCCTG---------GATGAGTCAC
Manatee	AGTGATGGCCTG---------GATGACTTGC
Asian elephant	AGTGACGGCCTG---------GATGTCTTAA
Rock hyrax	AGTGACAACCTA---------AGTGATTCAC
Sloth	AGTGATGACATACTAACTTCTGATGACTCAC
Flying lemur	AGTGATGAAATTTTAGCTTCTGATGACTCAC
Jackrabbit	AGTAATGAAATGTTAACTCCTGATGACTCAC
Flying squirrel	AGTGACGAAATGTTAACTTCTGATGACTCAG
Mouse	ACTGGTGAAATGTTAACTTCTGACAGCGCAT
Tree shrew	AGTGATGAAATGTTAACTTCTAACGACTCAC
Human	AGTGATGAACTGTTAGGTTCTGATGACTCAC
Hedgehog	AGTGATGAACTGTTAACTTCTGATGACTCAT
Flying fox	CGTGATGAAATATTAACTTCTGATGTCTCAC
Little brown bat	CGTGATGAAATACTAACTTCTGATGGCTCAC
Pangolin	AGTGATGAAATGTTAACTTCTGATGATCCAT
Dog	AGTGACGAAATATTAACTTCTGATGATTCAC
Horse	AGTGAGGAAATGTTAACTTCTGATGACTCAT
Cow	AGTGATGAAATATTAACTTCTGATGACTCGT
Humpback whale	AGTGATGAAATGTTAACTTCTAACGACTCAC

Figure 2 Positions 353–383 of the BRCA1 nucleotide alignment for representative taxa to show the 9-bp deletion that occurs in all 12 afrotherians but not in 39 other placental mammals or in the 3 marsupials that were investigated.

Methods

Data collection and taxon sampling

Alpha 2B adrenergic receptor (A2AB) gene sequences (AJ251174–AJ251187) were obtained as described elsewhere[5] except for *Bradypus tridactylus*, where betaine (1 M) and dimethyl sulphoxide (1.3%) were added for polymerase chain reaction (PCR) amplification. 12S rRNA, transfer RNA valine and 16S rRNA sequences were obtained for *Otolemur crassicaudatus* (AF179289), and a 16S rRNA sequence was obtained for *Dugong dugon* (AF179291) as described[5]. Additional sequences for A2AB, 12S rRNA, tRNA valine, and 16S rRNA were extracted from GenBank, as were sequences for the nuclear interphotoreceptor retinoid binding protein (IRBP) and von Willebrand Factor (vWF) genes. Species names and accession numbers are available in the Supplementary Information. Thirty-two new placental *BRCA1* sequences (AF283999–AF284030) were amplified using described primers[7] or, in a few cases, with additional primers. Among marsupials, a 1.0-kb fragment of *Macropus BRCA1* (AF284033) was amplified using BRCA1F26 (ref. 7) and R1164 (5′- tagargacttcytcctcagcct-3′), a 0.7-kb fragment of *Lutreolina* BRCA1 (AF284032) was amplified using F498 (5′-gaaagttaatgagtggttttccagaa-3′) and R1164, and overlapping fragments of *Vombatus BRCA1* (AF284031) were amplified using MBF1 (5′-gagccatgtggcacarayaytc-3′) with MBR6 (5′- tcagtcttcyca-gaaaygctagaatttc-3′), F498 with R1164 and BRCVom1 (5′-agcttgaattaaatagtagaa-3′) with R3042BR (5′-aagtgttggaagcagggaagctcttc-3′). Combing data was done as described[7].

Phylogenetic analysis

Sequences were aligned using CLUSTAL W[23]. Ambiguous regions of the 12S rRNA/tRNA valine-16S rRNA alignment were excluded from phylogenetic analyses as were a glutamic

acid repeat region of the A2AB gene and a 21-bp region of the *BRCA1* gene that is repeated up to four times. The aligned data sets were the following lengths: A2AB (1,164 bp); IRBP (1,292 bp); vWF (1,251 bp); 12S rRNA/tRNA valine-16S rRNA (2,001 bp); and BRCA1 (2,947 bp). Alignments for the concatenated (A2AB + IRBP + vWF + rRNA) and BRCA1 data sets are available in the Supplementary Information. Phylogenetic analyses included unweighted and transversion parsimony, minimum evolution (5,708-bp data set) or neighbour joining (BRCA1 data set) with logdet and maximum likelihood-GTR[24] distances, neighbour joining with weighted average (WAVE) maximum likelihood distances[25], and maximum likelihood under the HKY85 (ref. 24) model of sequence evolution. Gaps were coded as missing in parsimony analyses. Maximum-likelihood estimates of relative rates and transition to transversion ratios were obtained from maximum parsimony trees and used in subsequent maximum-likelihood analyses and in calculating weighted-average maximum-likelihood distances. Bootstrap support values are based on 500 replications except for maximum likelihood (100 replications). Maximum-likelihood bootstrap analyses with the BRCA1 data set used the following backbone constraint, where taxon numbers correspond to the ordering of taxa (top to bottom) in Fig. 1B: (((1–4), (5, 6), (7, 8)), ((9, 10), 11-14), (15, 16), (17, 18), 19–21, ((((22–24), 25), 26), 27), 28, (29, 30), 31–35, ((36, 37), (38, 39), (40, 41)), ((42, 43), 44), (45, 46), 47, ((48, 49), 50, 51)). Kishino and Hasegawa tests[24] were used to examine *a priori* hypotheses and to examine statistically acceptable root locations. In the latter case, we obtained the best unrooted likelihood tree for each data set and then evaluated all possible root positions. All phylogenetic analyses and statistical tests were performed with PAUP 4.0b2 (ref. 26), except for neighbour-joining with weighted average distances, where analyses were performed with PHYLIP 3.572 (J. Felsenstein) and WAVEBOOT (D. King and C. Krajewski). Maximum-likelihood analyses with rate partitions allowed the following eight rate partitions with the 5,708-bp data set: third positions of each nuclear gene; first + second positions of each nuclear gene; RNA stems; and RNA loops. Two rate partitions, corresponding to first + second and third codon positions, respectively, were used with the BRCA1 data set. NJ-WAVE analyses used a weighted-average distance approach[25] with the eight partitions indicated above for the 5,708-bp data set and two partitions (first + second codon positions; third codon positions) for the 2,947-bp data set; each partition was allowed its own rate, base composition, and transition to transversion ratio. Monte Carlo simulations were performed with Seq-Gen 1.1 (ref. 27) (Supplementary Information). Molecular dates were estimated using QDATE[28] (Supplementary Information).

Received 30 August; accepted 10 October 2000.

1. Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356,** 121–125 (1992).
2. Shoshani, J. & McKenna, M. C. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. *Mol. Phylogenet. Evol.* **9,** 572–584 (1998).
3. de Jong, W. W. Molecules remodel the mammalian tree. *Trends Ecol. Evol.* **13,** 270–275 (1998).
4. Springer, M. S., Burk, A., Kavanagh, J. R., Waddell, V. G. & Stanhope, M. J. The interphotoreceptor retinoid binding protein gene in therian mammals: implications for higher level relationships and evidence for loss of function in the marsupial mole. *Proc. Natl Acad. Sci. USA* **94,** 13754–13759 (1997).
5. Springer, M. S. *et al.* Endemic African mammals shake the phylogenetic tree. *Nature* **388,** 61–64 (1997).
6. Stanhope, M. J. *et al.* Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl Acad. Sci. USA* **95,** 9967–9972 (1998).
7. Teeling, E. C. *et al.* Molecular evidence regarding the origin of echolocation and flight in bats. *Nature* **403,** 188–192 (2000).
8. Asher, R. J. A morphological basis for assessing the phylogeny of the 'Tenrecoidea' (Mammalia, Lipotyphla). *Cladistics* **15,** 231–252 (1999).
9. Luckett, W. P. & Hartenberger, J.-L. Monophyly or polyphyly of the order Rodentia: Possible conflict between morphological and molecular interpretations. *J. Mammal. Evol.* **1,** 127–147 (1993).
10. Waddell, P. J., Cao, Y., Hauf, J. & Hasegawa, M. Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites-logdet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo and elephant. *Syst. Biol.* **48,** 31–53 (1999).
11. Penny, D., Masegawa, M., Waddell, P. J. & Hendy, M. D. Mammalian evolution: Timing and implications from using the logdeterminant transform for proteins of differing amino acid composition. *Syst. Biol.* **48,** 76–93 (1999).
12. McKenna, M. C. & Bell, S. K. *Classification of Mammals Above the Species Level* (Columbia Univ. Press, New York, 1997).
13. Rainger, R. *Agenda for Antiquity: Henry Fairfield Osborn and Vertebrate Paleontology at the American Museum of Natural History, 1890–1935* (Univ. Alabama Press, Tuscaloosa, Alabama, 1991).
14. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392,** 917–920 (1998).
15. Foote, M., Hunter, J. P., Janis, C. M. & Sepkoski, J. J. Jr Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science* **283,** 1310–1314 (1999).
16. Rich, T. H. *et al.* A tribosphenic mammal from the Mesozoic of Australia. *Science* **278,** 1438–1442 (1997).
17. Mouchaty, S. K., Gullberg, A., Janke, A. & Arnason, U. The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. *Mol. Biol. Evol.* **17,** 60–67 (2000).
18. Waddell, P. J., Okada, N. & Hasegawa, M. Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48,** 1–5 (1999).
19. Rasmussen, D. T. in *The Evolution of Perissodactyls* (eds Prothero, D. R. & Schoch, R. M.) 57–78 (Oxford Univ. Press, Oxford, 1989).
20. Matthew, W. D. The Carnivora and Insectivora of the Bridger basin, middle Eocene. *Mem. Am. Nat. Hist.* **9,** 291–567 (1909).
21. Novacek, M. J. The skull of leptictid insectivorans and the higher-level classification of eutherian mammals. *Bull. Am. Mus. Nat. Hist.* **183,** 1–111 (1986).
22. Easteal, S. Molecular evidence for the early divergence of placental mammals. *BioEssays* **21,** 1052–1058 (1999).
23. Thompson, J. D., Higgins, G. D. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).
24. Swofford, D. L., Olsen, G. P., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–492 (Sinauer, Sunderland, Massachusetts, 1996).
25. Krajewski, C., Blacket, M., Buckley, L. & Westerman, M. A multigene assessment of phylogenetic relationships within the dasyurid marsupial subfamily Sminthopsinae. *Mol. Phylogenet. Evol.* **8,** 236–248 (1997).
26. Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods) Version 4, (Sinauer, Sunderland, Massachusetts, 1998).
27. Rambaut, A. & Grassly, N. C. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13,** 303–306 (1997).
28. Rambaut, A. & Bromham, L. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15,** 442–448 (1998).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank F. Catzeflis for tissue samples. This work was supported by the NSF (M.S.S.) and the TMR program of the European Commission (W.W.d.J.; M.J.S.).

Chapter Seven

Molecular evolution of the alpha 2B adrenergic receptor

Ole Madsen, Diederik Willemsen, Björn M. Ursing, Ulfur Arnason and Wilfried W. de Jong. (*In press*) Molecular Evolution of the Mammalian Alpha 2B Adrenergic Receptor. *Mol. Biol. Evol.*

The alpha 2B adrenergic receptor (A2AB) is a heptahelical G protein–coupled receptor for catecholamines. We compared the almost complete coding region (about 1,175 bp) of the A2AB gene from 48 mammalian species, including eight newly determined sequences, representing all the 18 eutherian and two marsupial orders. Comparison of the encoded proteins reveals that residues thought to be involved in agonist binding are highly conserved, as are the regions playing a role in G protein–coupling. The three extracellular loops are generally more variable than the transmembrane domains and two of the intracellular loops, indicating a lower functional constraint. However, the greatest variation is observed in the very long, third intracellular loop, where only a few residues and a polyglutamyl tract are preserved. Although this polyglutamyl domain displays a great variation in length, its presence in all described A2ABs confirms its proposed role in agonist-dependent phosphorylation of the third intracellular loop. Phylogenetic analyses of the A2AB data set, including Bayesian methods, recognized the superordinal clades Afrotheria, Laurasiatheria, and Euarchontoglires, in agreement with recent molecular evidence, albeit with lower support. Within Afrotheria, A2AB strongly supports the paenungulate clade and the association of the continental African otter shrew with Malagasy tenrecs. Among Laurasiatheria, A2AB confirms the nesting of whales within the artiodactyls, as a sister group to hippopotamus. Within the Euarchontoglires, there is constant support for rodent monophyly.

Introduction

The gene for the alpha 2B adrenergic receptor (SWISS-PROT name A2AB) has extensively been used in recent studies on mammalian phylogeny (Springer et al. 1997; Stanhope et al. 1998; Madsen et al. 2001; Murphy et al. 2001*b*). A2AB has a length of about 450 amino acid residues and is encoded by a single copy gene (Lomasney et al. 1990). The gene is intronless, so the coding sequence can almost completely be amplified and sequenced from genomic DNA. Thus, an analysis of the deduced protein sequences can provide a fairly complete impression of the evolution of the protein.

A2AB is a representative of the biomedically highly important superfamily of G protein–coupled receptors (GPCRs) (Strader et al. 1995). With at least 616 members in the human genome, it is one of the largest known gene families (Venter et al. 2001) and can be divided into six groups: class A to E and the frizzled-smoothened family (Bockaert and Pin 1999; Horn, Vriend, and Cohen 2001; see http://www.gpcr.org/7tm/). These membrane receptors mediate signals induced by extracellular stimuli to the intracellular environment by way of guanyl nucleotide–binding proteins (G proteins). GPCRs are characterized by a bundle of seven transmembrane (TM) helices (TM1–7), which are connected by three extra- and three intracellular loops (EL1–3 and IL1–3, respectively), and have an extracellular N-terminus and an intracellular C-terminus (Baldwin 1993). The only known GPCR crystal structure is that of rhodopsin, which is a class-A receptor (Palczewski et al. 2000). A2AB is also a class-A receptor and is one of the adrenoceptors (adrenergic receptors) in the subfamily of amine receptors. Three types of mammalian adrenoceptors can be distinguished: alpha 1, alpha 2, and beta, each again having at least three subtypes. Class-A receptors bind mostly small ligands inside the helical bundle and are characterized by a conserved aspartate in TM2, important for monovalent cation regulation, and a tripeptide DRY or ERW at the intracellular end of TM3, involved in G-protein coupling (Ceresa and Limbird 1994; Oliveira et al. 1994; Strader et al. 1995; Scheer et al. 1996).

A2AB is a receptor for catecholamines, such as adrenaline and nonadrenaline, and mainly mediates the inhibition of adenylyl cyclase by way of coupling to $G_{i/o}$-proteins (Limbird 1988), but stimulation of adenylyl cyclase by way of coupling to G_s-proteins has also been reported (Eason et al. 1992). A2AB has a distinct expression in various tissues, mostly in the periphery, with highest expression in the kidney (e.g., Eason and Liggett 1993; Link et al. 1996). Knockout studies in mouse have shown that lack of the A2AB gene influences viability, the response to salt-induced hypertension, and blood pressure responses to agonists (Link et al. 1996; Makaritsis et al. 2000). Special features of A2AB are the absence of N-linked glycosylation sites in the extracellular domains and the presence of a very long, third intracellular loop, which contains a unique glutamate repeat. The third intracellular loop is necessary for microtubule sorting of A2AB to the cell surface (Saunders and Limbird 2000), whereas the polyglutamate domain is required for short-term agonist-promoted phosphorylation and consequent desensitization of A2AB (Jewell-Motz and Liggett 1995; Small et al. 2001). As with

Table 1
Species Names and Corresponding Accession Numbers of Sequences Used in this Study

Scientific name	Common Name	Accession Number	Scientific Name	Common Name	Accession Number
Amblysomus hottentotus	Golden mole	Y12526	*Macrotus californicus*	Leaf-nosed bat	AJ251180
Balaenoptera physalus	Fin whale	AJ251175	*Manis* sp.	Asian pangolin	AJ251185
Bos taurus	Bovine	Y15944	*Manis tetradactyla*	Long-tailed pangolin	AJ505820*
Bradypus tridactylus	3-toed sloth	AJ251179	*Megaderma lyra*	Vampire bat	AF337537
Cavia porcellus	Guinea pig	AJ271336*	*Micropotamogale lamottei*	Otter shrew	AJ251107*
Cavia porcellus	Guinea pig	U25723	*Mus musculus*	Mouse	L00979
Cynocephalus variegatus	Flying lemur	AJ251182	*Myotis daubentoni*	Daubenton's bat	AF337540
Cynopterus sphinx	Fruit bat	AJ251181	*Nycticebus coucang*	Loris	AJ251186
Diceros bicornis	Black rhino	AJ251184	*Orycteropus afer*	Aardvark	Y12522
Didelphis marsupialis	Opossum	Y15943	*Oryctolagus cuniculus*	Rabbit	Y15946
Dugong dugon	Dugong	Y15947	*Phoca vitulina*	Seal	AJ251176
Echinops telfairi	Madag hedgehog	Y17692	*Physeter catodon*	Sperm whale	AJ427417*
Elephas maximus	Asian elephant	Y12525	*Procavia capensis*	Hyrax	Y12523
Equus caballus	Horse	Y15945	*Pteropus rayneri*	Flying fox	AF337539
Erinaceus europaeus	Hedgehog	Y12521	*Rattus norvegicus*	Rat	M32061
Felis catus	Cat	AJ251174	*Sciurus vulgaris*	Squirrel	AJ315942
Hippopotamus amphibius	Hippo	AJ251178	*Sus scrofa*	Pig	AJ251177
Hipposideros commersoni	Leaf-nosed bat	AF337538	*Tadarida brasiliensis*	Free-tailed bat	AF337542
Homo sapiens	Human	M38742	*Talpa europaea*	Mole	Y12520
Homo sapiens	Human	AF316895	*Taphozous* sp.	Sheath-tailed bat	AF337543
Hylomys suillus	Gymnure	AJ505819*	*Tapirus terrestris*	Tapir	AJ315939
Lagenorhynchus albirostris	Dolphin	AJ315940	*Tenrec ecaudatus*	Common tenrec	AJ251108*
Lama pacos	Alpaca	AJ315941	*Tonatia bidens*	Round-eared bat	AF337541
Lama pacos	Alpaca	AJ505821*	*Trichechus manatus*	Manatee	AJ251109*
Macropus rufus	Kangaroo	AJ251183	*Tupaia tana*	Tree shrew	AJ251187
Macroscelides proboscideus	Elephant shrew	Y12524			

* Indicates new sequences for this study.

many other GPCRs (Pierce and Lefkowitz 2001), desensitization and internalization of A2AB is mediated by arrestins (DeGraff et al. 1999).

Little attention has generally been given to the molecular evolutionary information of genes used in phylogenetic studies. In this article we analyze the evolution of the structure and function of A2AB from 48 mammalian species. In addition, we also use a subset of 41 A2AB sequences to compare the phylogenetic relationships supported by the A2AB sequences with those obtained from concatenated data sets (e.g., Madsen et al. 2001; Murphy et al. 2001*b*). Combined data sets are required to get a robust phylogenetic tree, but it remains important to know the congruence with results obtained from single genes.

Materials and Methods

Eight new A2AB sequences are reported in this study and combined with 43 A2AB sequences extracted from the EMBL databank (release 68, September 2001) (table 1). The new A2AB sequences were obtained as described elsewhere (Springer et al. 1997; Madsen et al. 2001). In short, the primers A2ABFOR (5′-ascccta-ctcngtgcaggcnacng-3′) and A2ABREV (5′-ctgttgcagta-gccdatccaraaraaraaytg-3′) were used for PCR amplification on genomic DNA with the Expand High Fidelity PCR system (Boehringer Mannheim). PCR products were directly sequenced on both strands with internal primers or cloned into a pGEM-T vector (Promega), or both. If a sequence was determined from cloned DNA only, clones from at least two independent PCRs were sequenced to detect ambiguity caused by the PCR or allelic variations, or both. Sequencing was done with the Thermo Sequenase fluorescent-labeled primer or Thermo Sequenase Cy5 Dye Terminator cycle sequencing kits (Amersham Pharmacia Biotech).

Amino acid sequence alignments were made with GCG PILEUP (Wisconsin Package Version 10.0, GCG, Madison) and further optimized by eye. The nucleotide alignment was adjusted to correspond with the amino acid alignment.

To reduce computing time, seven of the nine bat species were excluded from the phylogenetic analyses because the relationships within bats (as based on A2AB) has recently been shown (Springer et al. 2001). Furthermore, we only used one of the alleles from human, alpaca, and guinea pig. Phylogenetic analyses were done on a 39-taxon (eutherians only) and a 41-taxon data set (eutherians plus out group marsupials) (see *Results*). Because great variability in base composition is observed at third codon positions (data not shown), phylogenetic analyses were performed with third codon positions unweighted and as transversions only, and without third codon positions. We further used the logdet model of sequence evolution (e.g., Lake 1994) in distance analyses to compensate for any possible heterogeneity in base composition among species. The glutamic acid domain in IL3 was excluded in phylogenetic analyses because it is difficult to align.

The following methods were used to obtain phylogenetic trees: maximum parsimony on DNA and amino acids, minimum evolution on DNA with logdet distances (e.g., Lake 1994), bio–neighbor joining (Gascuel 1997) on protein with JTT matrix distances (Jones,

Taylor, and Thornton 1992), maximum likelihood on DNA, and Bayesian posterior probability on DNA. PAUP4.0b2–4 (Swofford 1998) was used for parsimony, distance on DNA, and maximum likelihood analyses. The programs SEQBOOT, PROTDIST, and CONSENSE from the PHYLIP package (Felsenstein 2001), and BIONJ (Gascuel 1997) were used for distance analyses on amino acids.

Parsimony analyses included first, second, and third codon position unweighted; first and second position unweighted, with third position transversions only; first and second position unweighted; and amino acid unweighted. Ten random input orders of sequences were used, with gaps scored as missing, and in all PAUP analyses, the tree bisection–reconnection branch swapping option was used to swap branches. Bootstrap analyses included 500 or 100 replicates for DNA (parsimony-distance and maximum likelihood, respectively) and 250 replicates for amino acid sequences.

Maximum likelihood analyses were done with the HKY85 (Hasegawa, Kishino, and Yano 1985) model of sequence evolution. Transition to transversion ratios were calculated on minimum evolution trees obtained with HKY85 distances. These values were subsequently used for maximum likelihood calculations.

Bayesian phylogenetic analyses were performed with MRBAYES 2.1 (Huelsenbeck and Ronquist 2001). First, modeltest 3.06 (Posada and Crandall 1998) was used to determine which model of sequence evolution best fits the data under the maximum likelihood assumption. The best ML model of sequence evolution was subsequently used in Bayesian analyses, and the Metropolis-coupled Markov chain Monte Carlo sampling approach was used to calculate posterior probabilities. Initial probabilities for all trees were equal, and starting trees were random. Four Markov chains were run simultaneously 200,000, 350,000, and 500,000 times (to check consistency of results), tree sampling was done every 10 generations, and burn-in values were determined from the likelihood values.

Partitioned phylogenetic analyses included the following four partitions: extracellular (positions 14–24, 61–97, 138–181, and 395–422 in fig. 1), intracellular (25–60, 98–137, 182–298, and 320–394), intracellular without the variable parts of IL3 (25–60, 98–137, 182–202, and 372–394), and the variable parts of IL3 (203–298; and 320–371). Analyses on partitioned data were on protein with JTT matrix distances only, and the number of bootstrap replicates were 100 (see above for details).

The rate of nonsynonymous nucleotide substitutions (NNS) per site between human and mouse A2AB was calculated with MEGA 2.0 (Kumar et al. 2001), using the method of Li, Wu, and Luo (1985). The divergence between human and mouse was set to 80 Myr (to make NNS of A2AB comparable with that of Graur and Li 2000), and the rate unit is per site per billion years. This unit was used to facilitate the comparison of the NNS rate of A2AB with that of other mammalian protein-coding genes. The part of the human and mouse A2AB used to calculate the NNS corresponds to positions 14–405 in the human sequence, and the glutamic acid domain in IL3 was included in this calculation.

To detect whether the A2AB data set contains any significantly long branches, we performed RASA analyses using the program RASA 3.01 (Lyons-Weiler 2001).

Results and Discussion

Molecular Evolution of A2AB

A2AB sequences, including the eight newly determined ones, are now available for 48 mammalian species. These represent all the 18 eutherian orders and the major subordinal groups, as well as two species from divergent marsupial orders. Figure 1 relates the amino acid sequence variation deduced from the amplified part of the A2AB genes to the structural features of the protein. This composite sequence immediately reveals that most residues in the TM domains, in the intracellular loops IL1 and IL2, and in the extracellular loop EL1 are strictly conserved. It also emphasizes the huge sequence variability in the long IL3 and in EL2, and less so in EL3. In IL3, as many as nine different residues can be found at corresponding positions in different A2AB sequences, and it is the only region where indels occur. The observed sequence conservation and variability can be interpreted as follows.

Transmembrane Domains

Residues in TM3, TM5, and TM6 that are thought to be involved in ligand binding (* in fig. 1) are all conserved. The same is true for the intracellular ends of TM3 and TM6 that participate in G-protein coupling (% in fig. 1). The distribution of variable positions in the TM domains is interesting. Residues facing inward into the TM bundle are better conserved than those facing outward to the lipid bilayer (black and gray, respectively, in fig. 1). Of the former, only 9% are variable, against 34% of the lipid-facing positions, which also allow multiple replacements more often. Six of the seven inward-facing replacements are conserved with respect to side chain size, whereas residues oriented toward the lipid layer show greater size variation. Inward-facing residues are clearly more constrained than outward-facing ones in terms of size and chemical properties, reflecting their closer packing and tighter interaction with other residues in the TM bundle.

Extracellular Loops

Despite considerable sequence variation, it is noteworthy that all three extracellular loops have conserved their lengths. This suggests that the length of these loops is important for maintaining the proper conformation of the receptor. In EL1, only one position (81) is variable, whereas the other five form a conserved "GYWYF motif" (positions 75–80). The sequence YWYF is also retained in other alpha 2 adrenergic receptors (A2AA and A2AC), whereas W^{78} is conserved in all amine receptors but not in other class-A (rhodopsin-like) receptors (http://www.gpcr.org/7tm/). In the crystal structure of

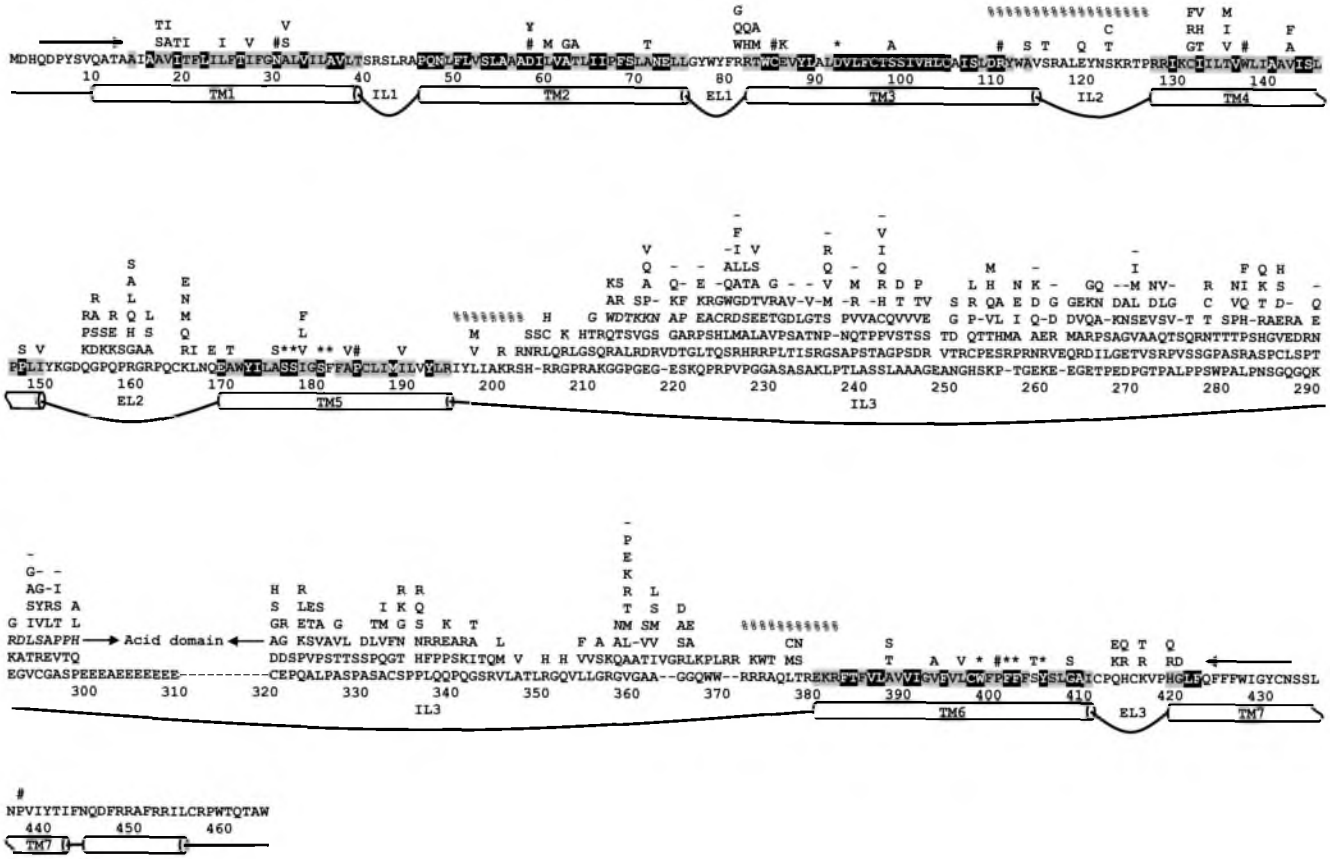


FIG. 1.—Sequence variation in mammalian A2AB. The observed variation is given above the majority consensus sequence (positions 14–422), as deduced from an alignment of A2AB sequences from 46 eutherian and two marsupial species. Because of the localization of the PCR primers, all but four of the 48 sequences miss the nucleotides coding for the 13 N-terminal and 42 C-terminal residues of the protein. The sequence from positions 1 to 13 and from 423 to 464 is human and includes the positions of the two primers used for PCR (arrows). Gaps are indicated by –. Residues involved in ligand binding in other adrenergic receptors (B2AR, A1AB, and A2AA; SWISS-PROT names (Ballesteros, Shi, and Javitch 2001)) are indicated by *, and ¥ indicates the conserved D^{58} important in monovalent cation regulation of GPCRs (Ceresa and Limbird 1994). G-protein coupling regions predicted in the three alpha 2 adrenoreceptor subtypes A2AA, A2AB, and A2AC (Wade et al. 1994; Eason and Liggett 1995; Saunders and Limbird 2000) are indicated by %. The secondary structure of A2AB, as predicted from the crystal structure of bovine rhodopsin (Palczewski et al. 2000), shows the positions of TM regions (TM1–7), intracellular loops (IL1–3), and extracellular loops (EL1–3). To predict the secondary structure, highly conserved TM positions (#) were used to align A2AB with bovine rhodopsin (Ballesteros, Shi, and Javitch 2001). Side chains in the TM domains predicted to face the lipid bilayer are marked in gray; those facing inward in the TM bundle are in black; and those unassigned are in white. For the glutamic acid domain in IL3, only the consensus sequence is given, and the variability of this domain in different species is presented in figure 2.

rhodopsin, EL1 as well as EL3 run along the periphery of the receptor (Palczewski et al. 2000). If this also applies to A2AB, EL1 may be in direct contact with other extracellular parts (EL2, EL3, distal parts of TM helices) thus stabilizing the conformation of the receptor. A functional role for W^{78} is supported by the finding that mutation to alanine or phenylalanine results in a 40-fold decrease in the binding of acetylcholine in the muscarinic m1 receptor (Matsui, Lazareno, and Birdsall 1995). Such a function of W^{78} would imply that EL1 is situated in the center of the receptor, close to or as a lid above the ligand-binding pocket, taking part in ligand binding or selection.

In EL2, 12 of the 19 residues are variable or even hypervariable. The conserved C^{164}, found in most GPCRs, forms a structurally important disulfide bond with C^{85} at the extracellular end of TM3 (Savarese, Wang, and Fraser 1992). The adjacent proline (P^{162}) might be conserved to allow EL2 to bend back into the helix bundle when forming the C^{85}–C^{164} disulfide bond. In rhodopsin, two antiparallel β-strands are present in EL2, corresponding to positions ~154–158 and ~163–166 in A2AB, of which the latter forms a part of the ligand-binding pocket (Palczewski et al. 2000). It seems unlikely that similar β-strands are present in A2AB because these positions are quite variable in this protein. The hypervariable nature of position 165 also makes it difficult to imagine that this region is a part of the binding-site crevice in A2AB. In rhodopsin, the residues corresponding to Y^{150} and N^{167} in A2AB are near each other, which, considering their conservation, may also be the case in A2AB. In class-A receptors, the N-terminal residues of EL2 are possibly important, either structurally or for ligand binding (Javitch et al. 2000). This could explain the relative conservation of this region in A2AB (positions 150–153).

Five out of the eight residues are conserved in EL3. Two of these, C^{415} and P^{418}, are retained in other alpha-2 adrenergic receptors, but not in the alpha-1 and beta-adrenergic receptor families, indicating a possible role

in the former (http://www.gpcr.org/7tm/). As mentioned, EL3 runs along the periphery of the receptor in rhodopsin. EL swapping experiments between beta-3 and alpha-1a adrenergic receptors have shown that EL3 plays a role in controlling receptor and G-protein affinity, probably by influencing the helical packing (Zhao, Gaivin, and Perez 1998). The same probably applies for EL3 in A2AB.

Intracellular Loops IL1 and IL2

The strict conservation of IL1 seems difficult to explain because no particular role has been suggested for this loop. Indeed, mutation in A1AB of the two basic residues in IL1 (corresponding to positions 41 and 44 in A2AB) does not have any effect on receptor activation and G-protein coupling (Greasley et al. 2001). On the other hand, IL1 in rhodopsin has a rigid organization (Palczewski et al. 2000) which may demand specific types of residues; the same may apply to A2AB, considering that the basic amino acids R^{41} (K in rhodopsin) and R^{44}, as well as L^{43} are conserved between rhodopsin and A2AB.

IL2 forms a part of the site for selective coupling of the G-protein complex (e.g., Ostrowski et al. 1992; Palczewski et al. 2000). Only three positions are conservatively variable, indicating that only few changes are tolerated to maintain subtype-specific G-protein coupling. One of these variable sites, S^{122}, may play a role together with R^{116} and A^{113} in the difference in dose-response–induced cAMP production between A2AB and A2AA in Sf9 cells (Nasman, Jansson, and Akerman 1997). Interestingly, although all placentals have A^{113}, this is S^{113} in the marsupial A2AB, as in the A2AA sequence at this position. This may suggest some differences in dose-response–induced cAMP production between placental and marsupial A2AB.

IL3 and the Polyglutamate Repeat

IL3 is extremely variable in A2AB, including many indels. However, the length is maintained, between 155 and 177 residues, and thus seems functionally important. In agreement with their role in G-protein coupling, the very N- and C-terminal ends of IL3 are relatively more conserved. In A2AA, both the N- and C-terminal ends are required for G_s coupling, whereas either one is sufficient for G_i coupling (Eason and Liggett 1996). The maintenance of two such G_i-coupling domains in one receptor is considered indicative of a strong evolutionary constraint. If A2AB also has redundant G_i-coupling domains, the greater conservation of the N- than the C-terminal end could reflect that they preferably couple different G_i proteins and therefore are under different evolutionary pressure to maintain optimal coupling.

In the hypervariable part of IL3, only very few residues and regions are conserved, notably E^{248}, P^{335}, positions ~346–~353, and the polyglutamate tract from position 299 to 319, shown in figure 1. All sequenced A2ABs possess such an acidic domain, but the length—and thus net acidity—varies considerably between species (fig. 2). The longest sequences are found in various

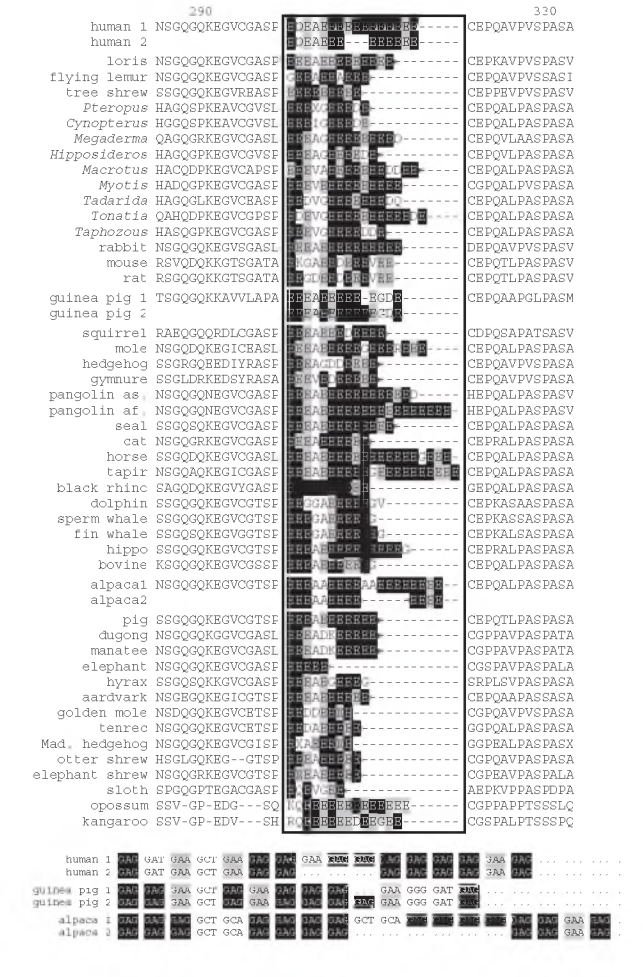


FIG. 2.—Glutamic acid domain in IL3 of 48 mammalian A2AB sequences. To better reveal the length variation, alignment of the polyE tracts is not optimized. Codon usage for glutamic acid in the polyE domains is indicated by black for GAG and gray for GAA. At the bottom, the allelic variation in human, guinea pig, and alpaca is given at the DNA level. For full species names, see *Materials and Methods.* Pangolin as. and pangolin af. refer to pangolin with Asian (*Manis* sp.) and African (*Manis tetradactyla*) origin, respectively.

cetferungulates (tapir, horse, pangolin, alpaca). Length polymorphisms within a species also occur. In humans, a three-residue deletion has been found (Baldwin et al. 1999; Heinonen et al. 1999; Small et al. 2001; Snapir et al. 2001), which has a frequency of 0.31 in Caucasians and 0.12 in African-Americans (Small et al. 2001). Among the 44 A2AB sequences that we determined, there was one case of length heterozygosity in the acid domain; in alpaca, two alleles were found, of which one had an 18 bp tandem repeat, coding for the sequence AAEEEE (alpaca 1, in fig. 2). Also, the guinea pig sample that we sequenced differed from the entry in the database in having one less Glu in the repeat (guinea pig 1, in fig. 2), in addition to five other amino acids differences.

Length variation of such poly–amino acid tracts is caused by unequal crossing-over and replication slippage, and is stabilized by point mutations (e.g., Smith 1976; Levinson and Gutman 1987; Jeffreys, Neil, and

Neumann 1998; Alba, Santibanez-Koref, and Hancock 1999). Length variation by replication slippage is characterized by identical codon repeats, whereas mixed codon usage may indicate unequal crossover, replication slippage followed by point mutations, or a combinations of these processes. As seen in figure 2, GAG (black) is the most frequently used Glu codon, but it is often interrupted by one or more GAA codons (gray) or codons for other amino acids. The latter are apparently under selective control, considering that next to E and D, almost exclusively A, V, and G seem to be allowed at positions 3–5 in the eutherian polyE tract. Whereas replication slippage is likely to occur in the GAG repeats, unequal crossing-over is more probably involved in the length polymorphisms in alpaca and humans (bottom, fig. 2).

Functional studies have shown that the acid domain is necessary for optimal agonist-promoted phosphorylation and the consequent desensitization of human A2AB (Jewell-Motz and Liggett 1995; Small et al. 2001). The shorter polyE allele in humans is associated with reduced metabolic rate in obese subjects and is a genetic risk factor for acute coronary events (Heinonen et al. 1999; Snapir et al. 2001). Also, in combination with a W64R polymorphism in the beta-3 adrenoceptor, an effect on fat mass is observed (Dionne et al. 2001). A full-length acid stretch is thus necessary for optimal receptor function of human A2AB. It has been suggested that the polyE tract provides the acid milieu that is required for the activity of the G protein–coupled receptor kinases (GRK) 2 and 3, which mediate the agonist-promoted phosphorylation of A2AB (Jewell-Motz and Liggett 1995). Such a functional role for acidic tracts might be more general because they are found in several other GPCRs (e.g., ACM1 and B1AR) and generally in their cytoplasmic domains. However, alignments of these GPCR sequences from different species (from http://www.gpcr.org/7tm/) reveal little length variations in their acid domains. Thus, in the case of A2AB, the conspicuous interspecies variation in length and acidity of the acid domain might possibly relate to the degree of phosphorylation needed in different species for desensitization and internalization of their A2AB receptors by way of beta-arrestin or reflect the conditions needed for optimal activity of GRKs in different species. A mutual adaptation between the level of desensitization needed for regulating activity, i.e., number of cell membrane receptors, and phosphorylation of different A2ABs can thus be envisaged (Daunt et al. 1997; DeGraff et al. 1999; Schramm and Limbird 1999).

IL3 is involved in microtubule-dependent sorting of A2AB (Saunders and Limbird 2000). The conservation of positions 346–353 may be important in this process to interact with proteins that bind to IL3, such as spinophilin and 14-3-3-ζ (Prezeau et al. 1999; Richman et al. 2001). Binding of arrestin to this domain is less likely because this is hyperphosphorylation dependent (Small et al. 2001), and there is only one phosphorylatable Thr in this domain. The conservation of E^{248} and P^{336} is difficult to explain. One might speculate that P^{336} allows IL3 to bend and bring the acid domain close to the serines and threonines that must be phosphorylated upon ligand binding and that E^{248} contributes to creating the proper acid milieu for GRK activity or is needed for phosphorylation of specific residues. Mutation studies are needed to confirm these speculations.

Evolutionary Rate of A2AB

Compared with other members of the adrenergic receptor family, A2AB seems to have one of the highest rates of evolutionary change. Whereas most adrenergic receptors, such as A1AB and B1AB, show identity scores around or above 90% between mouse-rat and humans (www.cmbi.kun.nl/7tm/), there is 82% identity in the case of A2AB. This is likely due to the great variability in the long IL3 in A2AB. Furthermore, the rate of NNS between mouse and human A2AB is 1.13 substitutions/site/billion years, which is in the higher range as compared with most other types of proteins (Graur and Li 2000, pp. 102–103). Insulin, for example, has a NNS rate of 0.20, myoglobin of 0.57, and prion protein of 0.29.

The A2AB Gene Tree

Different evolutionary constraints may act on different parts of a protein as a consequence of interactions with other cellular components (for GPCRs, see Donnelly, Findlay, and Blundell 1994). If this were the case for A2AB, different domains of the protein might provide different trees, which would be interesting from a functional point of view but would hamper the phylogenetic inferences. We therefore performed separate analyses on the extra- and intracellular domains of A2AB, with and without exclusion of the variable regions of IL3, and on this variable region itself (see *Materials and Methods*). Bootstrapped phylogenetic analyses of these four, partitioned A2AB data sets revealed no significantly supported differences in tree topologies, indicating that any possible distinct evolutionary constraints acting on the different partitions were not apparent in this manner (data not shown). A2AB was thus taken to evolve as a single genetic unit, making it useful for analysis of species phylogeny.

Phylogenetic analyses of concatenations of many genes, including A2AB, have been necessary to achieve a well-supported resolution of the eutherian tree (Murphy et al. 2001*b*). It nevertheless remains important to assess individual gene trees; strongly supported deviations from the combined tree might reveal peculiar mutational features in certain evolutionary lineages of that particular gene or might indicate the inclusion of paralogs. We therefore performed phylogenetic analyses on A2AB, both unrooted (eutherians only) and rooted (with marsupial out-groups). Unrooted analyses may detect relationships that are less pronounced in rooted analyses because out-group taxa may influence in-group relationships due to long branch attraction and sequence saturation (see Scally et al. 2001 for further explanation). The rooted maximum likelihood tree is shown in figure 3, and support values from various analyses are shown in table 2. Although certain lineages appear to evolve

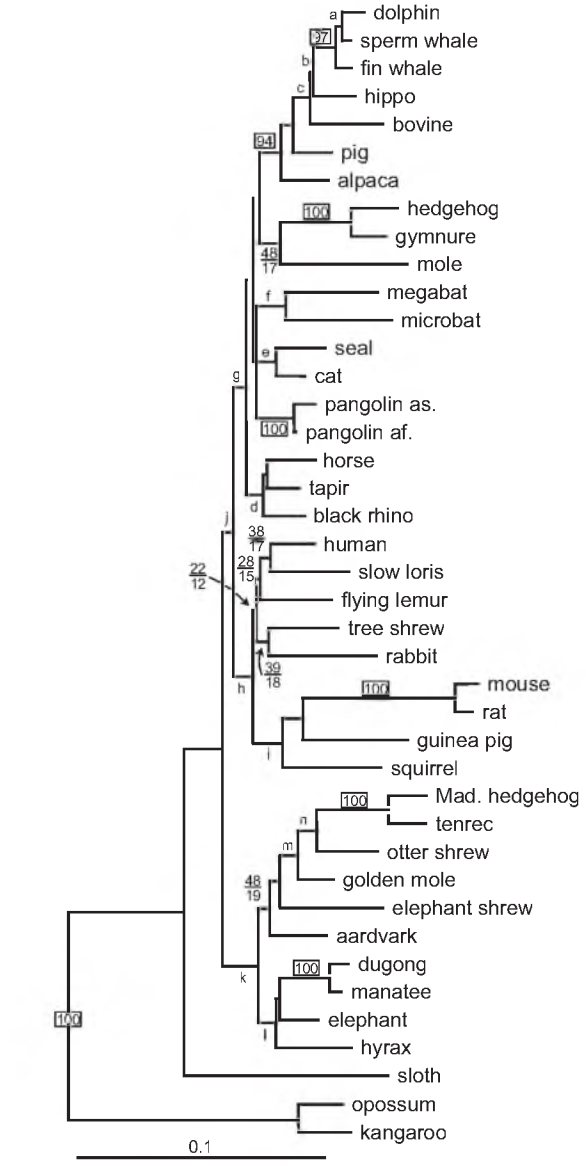


FIG. 3. Rooted maximum likelihood tree of eutherian A2AB sequences. The tree is based on first and second codon positions of a 41-taxon data set (−ln L = 6756.46) using the HKY85 model of sequence evolution, with transition to transversion ratio set to 1.46. See *Materials and Methods* for details. Branch lengths are proportional to the number of DNA changes, and the bar corresponds to 0.1 nucleotide change per site. Values in rectangles give average bootstrap-Bayesian support for nodes that are supported by more than 80% or 0.80, respectively, in all analyses described in table 2. Other nodes, supported by 50% or more in half of the analyses, are denoted a to n, and support values are shown in table 2. Nodes not supported by 50% or more in half of the analyses are marked with the average bootstrap-Bayesian support and the number of times that the node is supported (top and bottom values, respectively). Unmarked nodes are supported in less than half of the phylogenetic analyses. For species names, see *Materials and Methods*. Megabat is *Cynopterus sphinx* and microbat is *Macrotus californicus*. Pangolin as. and pangolin af. refer to pangolin with Asian (*Manis* sp.) and African (*Manis tetradactyla*) origin, respectively.

faster than others (fig. 3; e.g., murid rodents, tenrecs), RASA analyses showed that these branches are not significantly longer than the others.

In no instance does the A2AB tree deviate significantly from the recently emerged eutherian tree (Madsen et al. 2001; Murphy et al. 2001*a*, 2001*b*). It therefore seems that concerns of amplifying pseudogenes or paralogs by using an intronless gene from a large gene family are unfounded in the case of A2AB. Several nodes in the A2AB tree, however, are poorly resolved and supported. At the ordinal level, medium support is seen for Afrosoricida (average 73% in unrooted and 54% in rooted analyses), Rodentia (75%–72%), and Chiroptera (69%–68%), and weak support is seen for Primates (38%–38%) and Eulipotyphla (48%–49%). However, all analyses support rodent monophyly, varying from 54% in rooted unweighted parsimony to 1.00 in Bayesian probability analyses. This consistent support for monophyly is in line with all recent analyses of other nuclear genes (Huchon, Catzeflis, and Douzery 1999; Adkins et al. 2001; DeBry and Sagel 2001) and of combined nuclear and mitochondrial data sets (Madsen et al. 2001; Murphy et al. 2001*a*, 2001*b*) but contradicts analyses on total mitochondrial proteins that still indicate rodent para- or polyphyly (Reyes, Pesole, and Saccone 2000; Janke et al. 2002).

The division of lipotyphlan insectivores into two orders, Eulipotyphla and Afrosoricida, is supported by A2AB. Eulipotyphla is found in 17 of the 20 analyses but with low confidence values (fig. 3). However, we never observed a basal position for Erinaceidae (hedgehog and gymnure), as is found for whole mitochondrial DNA analyses (Krettek, Gullberg, and Arnason 1995; Mouchaty et al. 2000); Erinaceidae always groups inside Laurasiatheria, far from the eutherian root (see below). Support for Afrosoricida (golden moles and tenrecs) is mixed (branch m, fig. 3, table 2). Notably, there is an extreme difference in support between rooted and unrooted Bayesian probability analyses (0.01 and 0.99, respectively). This is caused by the placement of the root between the Malagasy tenrecs and continental African otter shrew in rooted Bayesian probability analyses (see below). Also, analyses which include unweighted third codon positions give low support for Afrosoricida. In most of these analyses, golden mole groups with aardvark due to third position transitions, whereas weighting as transversions-only restores Afrosoricida (data not shown). When compared with our previous analyses of A2AB (Stanhope et al. 1998), the increased taxon sampling of Afrosoricida has a positive influence on the support for this group. Within Afrosoricida, we observe a well-supported affiliation of otter shrew with the two Malagasy tenrecs (node n, mean support 94% and 80%). This confirms our prediction from sequence signatures that the otter shrew is a member of Afrotheria (van Dijk et al. 2001).

There is compelling molecular evidence, from nuclear and mitochondrial genes and from SINE elements, that whales are nested within Artiodactyla as sister to hippo's, and that ruminants are the sister group of the hippo-whale clade (e.g., Ursing and Arnason 1998; Ga-

Table 2
Bootstrap and Bayesian Probability Support for Branches a to n in Figure 3

	Parsimony								Distance								Bayesian		ML		Average	
	all		1+2		1+2+3tv		aa		all		1+2		1+2+3tv		aa		all		all		Sup.	
Node	39	41	39	41	39	41	39	41	39	41	39	41	39	41	39	41	39	41	39	41	39	41
a	96	96	76	76	76	79	70	68	56	58	62	60	62	57	84	81	98	97	95	91	78	76
b	83	83	46	46	30	30	29	29	98	98	80	77	87	86	66	69	99	99	94	95	71	71
c	48	43	60	59	59	54	76	72	58	56	62	57	64	69	50	53	88	86	72	62	64	61
d	96	97	65	67	77	75	25	32	100	100	90	90	94	95	69	73	100	100	92	92	81	82
e	99	99	84	85	98	99	70	70	100	100	79	82	96	98	50	58	100	100	99	98	88	89
f	81	75	70	71	81	78	64	57	76	74	46	46	42	44	54	49	100	100	74	85	69	68
g	85	81	28	29	66	63	13	18*	91	90	21	28	53	60	38	27	100	100	86	81	58	58
h	43	46	50	51	52	56	49	39	51	65	43	52	40	47	50	38	100	100	70	66	55	57
i	56	54	71	66	69	70	74	69	77	69	77	81	85	84	64	59	100	100	75	66	75	72
j	57	51	26	27	64	69	26	25	53	68	22	35	54	66	24	17	100	100	72	67	50	53
k	99	78	99	71	94	43	95	63	82	87	96	45	92	25	91	68	100	1	96	20*	94	50
l	91	85	69	62	90	90	64	64	98	99	79	82	94	89	91	92	100	100	96	74	87	84
m	18	19	98	87	96	58	86	81	23	22	95	93	94	87	73	62	99	1	50	26	73	54
n	99	98	96	88	99	65	99	91	97	96	76	72	99	97	71	74	100	31	99	82	94	80

Note.—Columns 39 and 41 refer to the 39- and 41-taxon data sets, for unrooted and rooted analyses, respectively (see *Materials and Methods*). All 1+2, 1+2+3tv, and aa refer to analyses on all codon positions, first plus second codon positions, first plus second plus transversions on third codon positions, and amino acids, respectively. Distance analyses on DNA and amino acids is based on logdet (e.g., Lake 1994) and JTT (Jones, Taylor, and Thornton 1992) matrix distances, respectively. Bayesian is the Bayesian posterior probability, as based on the GTR + Γ_8 + I (rate heterogeneity, gamma distribution of variable sites, and proportion invariable-sites) model of sequence evolution, four simultaneous Markov chains, 500,000 generations and burn-in value at 30,000 generations. Bayesian analyses based on 200,000 and 350,000 generations gave nearly identical results (maximum difference in probability of 0.02). Note that probability is shown as percentage (e.g., 97 corresponds to a probability of 0.97). ML is maximum likelihood analyses with rate heterogeneity (HKY85; Hasegawa, Kishino, and Yano 1985), where transition/transversion ratios were set to 1.80 and 1.73 for the 39- and 41-taxon data sets, respectively. * indicates that branch is not present in bootstrap/Bayesian majority rule consensus tree. Average sup. is the average support of the phylogenetic analyses.

tesy et al. 1999; Nikaido, Rooney, and Okada 1999). For A2AB, the mean support for the hippo-whale clade is 71% and 71% (unrooted-rooted), whereas ruminants (bovine) cluster with hippo-whale, with average support of 64% and 61% (nodes b and c; table 2). The position of pigs (Suiformes) and alpacas (Tylopoda) within Cetartiodactyla is unresolved by A2AB.

Recent analyses in which A2AB was combined with other genes strongly supported the division of Eutheria into four major groups (Madsen et al. 2001; Murphy et al. 2001*a*, 2001*b*). These four groups are also supported, albeit not conclusively, by A2AB itself: Afrotheria (branch k, mean support 94% and 50%), Laurasiatheria (branch g, mean 58% and 58%), Euarchontoglires (branch h, mean 55% and 57%), and Xenarthra (branch to sloth). In contrast to Laurasiatheria and Euarchontoglires, Afrotheria is strongly affected by rooting the tree, lowering the average support from 94% to 50%. Moreover, in the rooted analyses, the support ranges from 1% to 87%, caused by a tendency of the root to nest inside Afrotheria in some analyses (see table 2). The low support in rooted analyses, and the occasional rooting inside Afrotheria, indicates a basal position of Afrotheria in the A2AB tree, in accordance with the concatenated sequence evidence (Murphy et al. 2001*b*). It also illustrates the difficulty in establishing a stable root for the Eutherian tree in single gene analyses, an often neglected problem.

Joining of Laurasiatheria and Euarchontoglires in the Boreoeutheria (Springer et al. 2002) is reasonably well supported (branch j). Within the three superordinal clades, not much further resolution of orders is observed. The only relationship that is consistently supported with high bootstrap and probability values is Paenungulata within Afrotheria (node l, mean support 87% and 84%). Furthermore, it can be noted that most analyses support an aardvark-elephant shrew-Afrosoricida clade and rabbit as sister to tree shrew. The latter, however, is in contrast with the increasing molecular evidence for the grouping of rodents and lagomoprhs in Glires (e.g., Murphy et al. 2001*a*).

The inability of A2AB and other individual nuclear genes to robustly resolve various regions of the eutherian tree strongly advocates their concatenation in ever larger data sets. This may accumulate enough signal to ultimately resolve the remaining weakly supported and ambiguous relationships in the tree. Alternatively, one may hope that qualitative, rare genomic changes (Rokas and Holland 2000), like transposons and indels, may be helpful in this respect.

In conclusion, the comparison of A2AB sequences from a broad range of well-selected mammalian species has served a dual purpose. It has contributed—in combination with other genes—to the resolution of mammalian phylogeny and has additionally revealed the importance of sequence variation in the protein with respect to function and structure. Residues and regions involved in ligand binding and G-protein coupling are highly conserved, as are the major parts of the TM domains. Highest variation is found in the very long, third intracellular loop IL3, where only a few residues and a polyglutamyl domain are conserved. The presence of this polyglutamyl domain in all A2ABs, albeit with varying lengths, confirms its proposed role in agonist-dependent phosphorylation of IL3. Thus, comparing orthologous protein sequences from more mammalian species other than only humans, rat, and mouse is likely to

give valuable additional information about important parts of such a protein.

Acknowledgments

We thank Dr. Gert Vriend for critical reading of the manuscript, and Drs. François Catzeflis and Mark Springer for tissue and DNA samples. This work was supported by a TMR grant from the European Commission.

LITERATURE CITED

ADKINS, R. M., E. L. GELKE, D. ROWE, and R. L. HONEYCUTT. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. Mol. Biol. Evol. **18**:777–791.

ALBA, M. M., M. F. SANTIBANEZ-KOREF, and J. M. HANCOCK. 1999. Conservation of polyglutamine tract size between mice and humans depends on codon interruption. Mol. Biol. Evol. **16**:1641–1644.

BALDWIN, C. T., F. SCHWARTZ, J. BAIMA ET AL. (14 co-authors). 1999. Identification of a polymorphic glutamic acid stretch in the alpha2B-adrenergic receptor and lack of linkage with essential hypertension. Am. J. Hypertens. **12**:853–857.

BALDWIN, J. M. 1993. The probable arrangement of the helices in G protein-coupled receptors. EMBO J. **12**:1693–1703.

BALLESTEROS, J. A., L. SHI, and J. A. JAVITCH. 2001. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. Mol. Pharmacol. **60**: 1–19.

BOCKAERT, J., and J. P. PIN. 1999. Molecular tinkering of G protein-coupled receptors: an evolutionary success. EMBO J. **18**:1723–1729.

CERESA, B. P., and L. E. LIMBIRD. 1994. Mutation of an aspartate residue highly conserved among G protein-coupled receptors results in nonreciprocal disruption of alpha 2-adrenergic receptor-G-protein interactions. A negative charge at amino acid residue 79 forecasts alpha 2A-adrenergic receptor sensitivity to allosteric modulation by monovalent cations and fully effective receptor/G protein-coupling. J . Biol. Chem. **269:**29557–29564.

DAUNT, D. A., C. HURT, L. HEIN, J. KALLIO, F. FENG, and B. K. KOBILKA. 1997. Subtype-specific intracellular trafficking of alpha2-adrenergic receptors. Mol. Pharmacol. **51**:711–720.

DEBRY, R. W., and S. M. SAGEL. 2001. Phylogeny of rodentia (Mammalia) inferred from the nuclear-encoded gene IRBP. Mol. Phylogenet. Evol. **19**:290–301.

DEGRAFF, J. L., A. W. GAGNON, J. L. BENOVIC, and M. J. ORSINI. 1999. Role of arrestins in endocytosis and signaling of alpha2-adrenergic receptor subtypes. J. Biol. Chem. **274**: 11253–11259.

DIONNE, I. J., A. N. TURNER, A. TCHERNOF, T. I. POLLIN, D. AVRITHI, D. GRAY, A. R. SHULDINER, and E. T. POEHLMAN. 2001. Identification of an interactive effect of beta3- and alpha2b-adrenoceptor gene polymorphisms on fat mass in Caucasian women. Diabetes **50**:91–95.

DONNELLY, D., J. B. FINDLAY, and T. L. BLUNDELL. 1994. The evolution and structure of aminergic G protein-coupled receptors. Recept. Channels **2**:61–78.

EASON, M. G., H. KUROSE, B. D. HOLT, J. R. RAYMOND, and S. B. LIGGETT. 1992. Simultaneous coupling of alpha 2-adrenergic receptors to two G proteins with opposing effects. Subtype-selective coupling of alpha 2C10, alpha 2C4, and alpha 2C2 adrenergic receptors to Gi and Gs. J. Biol. Chem. **267**:15795–15801.

EASON, M. G., and S. B. LIGGETT. 1993. Human alpha 2-adrenergic receptor subtype distribution: widespread and subtype-selective expression of alpha 2C10, alpha 2C4, and alpha 2C2 mRNA in multiple tissues. Mol. Pharmacol. **44**: 70–75.

———. 1995. Identification of a Gs coupling domain in the amino terminus of the third intracellular loop of the alpha 2A-adrenergic receptor. Evidence for distinct structural determinants that confer Gs versus Gi coupling. J. Biol. Chem. **270**:24753–24760.

———. 1996. Chimeric mutagenesis of putative G protein-coupling domains of the alpha2A-adrenergic receptor. Localization of two redundant and fully competent Gi coupling domains. J. Biol. Chem. **271**:12826–12832.

FELSENSTEIN, J. 2001. PHYLIP (phylogeny inference package). Version 3.6a2.1. Department of Genetics, University of Washington, Seattle.

GASCUEL, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. **14**:685–695.

GATESY, J., M. MILINKOVITCH, V. WADDELL, and M. J. STANHOPE. 1999. Towards resolving the interordinal relationships of placental mammals. Syst. Biol. **48**:16–20.

GRAUR, D., and W. LI. 2000. Fundamentals of molecular evolution. Sinauer Associates, Inc., Mass.

GREASLEY, P. J., F. FANELLI, A. SCHEER, L. ABUIN, M. NENNIGER-TOSATO, P. G. DEBENEDETTI, and S. COTECCHIA. 2001. Mutational and computational analysis of the alpha(1b)-adrenergic receptor. Involvement of basic and hydrophobic residues in receptor activation and G protein coupling. J. Biol. Chem. **276**:46485–46494.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

HEINONEN, P., M. KOULU, U. PESONEN, M. K. KARVONEN, A. RISSANEN, M. LAAKSO, R. VALVE, M. UUSITUPA, and M. SCHEININ. 1999. Identification of a three-amino acid deletion in the alpha2B-adrenergic receptor that is associated with reduced basal metabolic rate in obese subjects. J. Clin. Endocrinol. Metab. **84**:2429–2433.

HORN VRIEND, F. G., and F. E. COHEN. 2001. Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems. Nucleic Acids Res. **29**:346–349.

HUCHON, D., F. M. CATZEFLIS, and E. J. DOUZERY. 1999. Molecular evolution of the nuclear von Willebrand factor gene in mammals and the phylogeny of rodents. Mol. Biol. Evol. **16**:577–589.

HUELSENBECK, J. P., and F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755.

JANKE, A., O. MAGNELL, G. WIECZORE, M. WESTERMAN, and U. ARNASON. 2002. Phylogenetic analysis of 18S rRNA and the mitochondrial genomes of the wombat, vombatus ursinus, and the spiny anteater, tachyglossus aculeatus: increased support for the marsupionta hypothesis. J. Mol. Evol. **54**:71–80.

JAVITCH, J. A., L. SHI, M. M. SIMPSON, J. CHEN, V. CHIAPPA, I. VISIERS, H. WEINSTEIN, and J. A. BALLESTEROS. 2000. The fourth transmembrane segment of the dopamine D2 receptor: accessibility in the binding-site crevice and position in the transmembrane bundle. Biochemistry **39**:12190–12199.

JEFFREYS, A. J., D. L. NEIL, and R. NEUMANN. 1998. Repeat instability at human minisatellites arising from meiotic recombination. EMBO J. **17**:4147–4157.

JEWELL-MOTZ, E. A., and S. B. LIGGETT. 1995. An acidic motif within the third intracellular loop of the alpha2C2 adrenergic receptor is required for agonist-promoted phosphorylation and desensitization. Biochemistry **34**:11946–11953.

JONES, D. T., W. R. TAYLOR, and J. M. THORNTON. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8**:275–282.

KRETTEK, A., A. GULLBERG, and U. ARNASON. 1995. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, Erinaceus europaeus, and the phylogenetic position of the Lipotyphla. J. Mol. Evol. **41**:952–957.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN, and M. NEI. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

LAKE, J. A. 1994. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. Proc. Natl. Acad. Sci. USA **15**:1455–1459.

LEVINSON, G., and G. A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.

LI, W. H., C. I. WU, and C. C. LUO. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.

LIMBIRD, L. E. 1988. Receptors linked to inhibition of adenylate cyclase: additional signaling mechanisms. FASEB J. **2**:2686–2695.

LINK, R. E., K. DESAI, L. HEIN, M. E. STEVENS, A. CHRUSCINSKI, D. BERNSTEIN, G. S. BARSH, and B. K. KOBILKA. 1996. Cardiovascular regulation in mice lacking alpha2-adrenergic receptor subtypes b and c. Science **273**:803–805.

LOMASNEY, J. W., W. LORENZ, L. F. ALLEN, K. KING, J. W. REGAN, T. L. YANG-FENG, M. G. CARON, and R. J. LEFKOWITZ. 1990. Expansion of the alpha 2-adrenergic receptor family: cloning and characterization of a human alpha 2-adrenergic receptor subtype, the gene for which is located on chromosome 2. Proc. Natl. Acad. Sci. USA **87**:5094–5098.

LYONS-WEILER, J. 2001. RASA 3.04 Turbo for the Macintosh. http://bio.uml.edu/LW/RASA.html.

MADSEN, O., M. SCALLY, C. J. DOUADY, D. J. KAO, R. W. DEBRY, R. ADKINS, H. M. AMRINE, M. J. STANHOPE, W. W. DE JONG, and M. S. SPRINGER. 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature **409**:610–614.

MAKARITSIS, K. P., C. JOHNS, I. GAVRAS, and H. GAVRAS. 2000. Role of alpha(2)-adrenergic receptor subtypes in the acute hypertensive response to hypertonic saline infusion in anephric mice. Hypertension **35**:609–613.

MATSUI, H., S. LAZARENO, and N. J. BIRDSALL. 1995. Probing of the location of the allosteric site on m1 muscarinic receptors by site-directed mutagenesis. Mol. Pharmacol. **47**: 88–98.

MOUCHATY, S. K., A. GULLBERG, A. JANKE, and U. ARNASON. 2000. The phylogenetic position of the Talpidae within eutheria based on analysis of complete mitochondrial sequences. Mol. Biol. Evol. **17**:60–67.

MURPHY, W. J., E. EIZIRIK, W. E. JOHNSON, Y. P. ZHANG, O. A. RYDER, and S. J. O'BRIEN. 2001*a*. Molecular phylogenetics and the origins of placental mammals. Nature **409**: 614–618.

MURPHY, W. J., E. EIZIRIK, S. J. O'BRIEN ET AL. (11 co-authors). 2001*b*. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science **294**:2348–2351.

NASMAN, J., C. C. JANSSON, and K. E. AKERMAN. 1997. The second intracellular loop of the alpha2-adrenergic receptors determines subtype-specific coupling to cAMP production. J. Biol. Chem. **272**:9703–9708.

NIKAIDO, M., A. P. ROONEY, and N. OKADA. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. USA **96**:10261–10266.

OLIVEIRA, L., A. C. PAIVA, C. SANDER, and G. VRIEND. 1994. A common step for signal transduction in G protein-coupled receptors. Trends Pharmacol. Sci. **15**:170–172.

OSTROWSKI, J., M. A. KJELSBERG, M. G. CARON, and R. J. LEFKOWITZ. 1992. Mutagenesis of the beta 2-adrenergic receptor: how structure elucidates function. Annu. Rev. Pharmacol. Toxicol. **32**:167–183.

PALCZEWSKI, K., T. KUMASAKA, T. HORI ET AL. (12 co-authors). 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. Science **289**:739–745.

PIERCE, K. L., and R. J. LEFKOWITZ. 2001. Classical and new roles of beta-arrestins in the regulation of G protein-coupled receptors. Nat. Rev. Neurosci. **2**:727–733.

POSADA, D., and K. A. CRANDALL. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817–818.

PREZEAU, L., J. G. RICHMAN, S. W. EDWARDS, and L. E. LIMBIRD. 1999. The zeta isoform of 14-3-3 proteins interacts with the third intracellular loop of different alpha2-adrenergic receptor subtypes. J. Biol. Chem. **274**:13462–13469.

REYES, A., G. PESOLE, and C. SACCONE. 2000. Long-branch attraction phenomenon and the impact of among-site rate variation on rodent phylogeny. Gene **259**:177–187.

RICHMAN, J. G., A. E. BRADY, Q. WANG, J. L. HENSEL, R. J. COLBRAN, and L. E. LIMBIRD. 2001. Agonist-regulated interaction between alpha2-adrenergic receptors and spinophilin. J. Biol. Chem. **276**:15003–15008.

ROKAS, A., and P. W. HOLLAND. 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. **15**:454–459.

SAUNDERS, C., and L. E. LIMBIRD. 2000. Microtubule-dependent regulation of alpha(2B) adrenergic receptors in polarized MDCKII cells requires the third intracellular loop but not G protein coupling. Mol. Pharmacol. **57**:44–52.

SAVARESE, T. M., C. D. WANG, and C. M. FRASER. 1992. Site-directed mutagenesis of the rat m1 muscarinic acetylcholine receptor. Role of conserved cysteines in receptor function. J. Biol. Chem. **267**:11439–11448.

SCALLY, M., O. MADSEN, C. J. DOUADY, W. W. DE JONG, M. J. STANHOPE, and M. S. SPRINGER. 2001. Molecular evidence for the major clades of placental mammals. J. Mamm. Evol. **8**:239–277.

SCHEER, A., F. FANELLI, T. COSTA, P. G. DE BENEDETTI, and S. COTECCHIA. 1996. Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. EMBO J. **15**:3566–3578.

SCHRAMM, N. L., and L. E. LIMBIRD. 1999. Stimulation of mitogen-activated protein kinase by G protein-coupled alpha(2)-adrenergic receptors does not require agonist-elicited endocytosis. J. Biol. Chem. **274**:24935–24940.

SMALL, K. M., K. M. BROWN, S. L. FORBES, and S. B. LIGGETT. 2001. Polymorphic deletion of three intracellular acidic residues of the alpha 2B-adrenergic receptor decreases G protein-coupled receptor kinase-mediated phosphorylation and desensitization. J. Biol. Chem. **267**:4917–4922.

SMITH, G. P. 1976. Evolution of repeated DNA sequences by unequal crossover. Science **191**:528–535.

SNAPIR, A., P. HEINONEN, T. P. TUOMAINEN ET AL. (14 co-authors). 2001. An insertion/deletion polymorphism in the alpha2B-adrenergic receptor gene is a novel genetic risk factor for acute coronary events. J. Am. Coll. Cardiol. **37**: 1516–1522.

SPRINGER, M. S., G. C. CLEVEN, O. MADSEN, W. W. DE JONG, V. G. WADDELL, H. M. AMRINE, and M. J. STANHOPE. 1997. Endemic African mammals shake the phylogenetic tree. Nature **388**:61–64.

SPRINGER, M. S., M. SCALLY, O. MADSEN, W. W. DE JONG, and M. J. STANHOPE. 2002. Dispersal and vicariance in the early evolutionary history of crown-group placentals. Syst. Biol. (in press).

SPRINGER, M. S., E. C. TEELING, O. MADSEN, M. J. STANHOPE, and W. W. DE JONG. 2001. Integrated fossil and molecular data reconstruct bat echolocation. Proc. Natl. Acad. Sci. USA **22**:6241–6246.

STANHOPE, M. J., V. G. WADDELL, O. MADSEN, W. W. DE JONG, S. B. HEDGES, G. C. CLEVEN, D. KAO, and M. S. SPRINGER. 1998. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. Proc. Natl. Acad. Sci. USA **95**:9967–9972.

STRADER, C. D., T. M. FONG, M. P. GRAZIANO, and M. R. TOTA. 1995. The family of G protein-coupled receptors. FASEB J. **9**:745–754.

SWOFFORD, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Mass.

URSING, B. M., and U. ARNASON. 1998. Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade. Proc. R. Soc. Lond. B: Biol. Sci. **265**:2251–2255.

VAN DIJK, M. A., O. MADSEN, F. CATZEFLIS, M. J. STANHOPE, W. W. DE JONG, and M. PAGEL. 2001. Protein sequence signatures support the African clade of mammals. Proc. Natl. Acad. Sci. USA **98**:188–193.

VENTER, J. C., M. D. ADAMS, E. W. MYERS ET AL. (274 co-authors). 2001. The sequence of the human genome. Science **291**:1304–1351 [published erratum appears in Science (2001) **292**:1838)].

WADE, S. M., H. M. DALMAN, S. Z. YANG, and R. R. NEUBIG. 1994. Multisite interactions of receptors and G proteins: enhanced potency of dimeric receptor peptides in modifying G protein function. Mol. Pharmacol. **45**:1191–1197.

ZHAO, M. M., R. J. GAIVIN, and D. M. PEREZ. 1998. The third extracellular loop of the beta2-adrenergic receptor can modulate receptor/G protein affinity. Mol. Pharmacol. **53**:524–529.

PEER BORK, reviewing editor

Accepted August 7, 2002

Chapter Eight

General Discussion - Summary

The aim of the present study was twofold. First, to gain a deeper knowledge about the phylogenetic relationships amongst extant mammals, as derived from extensive molecular data sets. Second, and simultaneously, to describe the molecular evolution of the nuclear genes used in these phylogenetic studies, in relation to their structures and functions.

Molecular Phylogeny

Resolution of the phylogenetic relationships of mammals in general and of placental (eutherian) mammals in particular, as based on morphological and molecular data, has proven difficult (de Jong 1998). In this thesis the gradual progress in establishing a well-resolved eutherian tree, by means of mostly nuclear-encoded genes, is described. The resulting tree, reflecting our best current estimate of the relationships between extant eutherian orders, is presented in figure 1. The silhouettes in this figure display the morphological variety of the 18 orders, and may facilitate the understanding of this chapter.

Early molecular studies dealing with eutherian phylogenetic relationships were often hampered by the use of only a few species, representing a limited number of the 18 traditionally recognized eutherian orders (e.g. Li et al. 1990; Cao et al. 1994; Honeycutt et al. 1995). Broadening the taxon sampling to include more orders and for different genes seemed to be the first step in getting a better molecular based eutherian tree. In chapter 2, the sequences of exon 1 of the AQP1 gene from 12 mammalian species, representing 12 different orders, were determined. Phylogenetic analyses were performed with these sequences alone and in combination with previously published αA-crystallin sequences. High support was found for grouping aardvark (Tubulidentata) and elephant shrew (Macroscelidea) with paenungulates (Hyracoidea, Sirenia and Proboscidea) in a so-called African clade which later was dubbed Afrotheria (chapter 5; see figure 1). This result fully conformed to studies by others on completely different nuclear genes like those for the interphotoreceptor retinoid-binding protein and for von Willebrand factor (Porter et al. 1996; Stanhope et al. 1996).
In the following chapters (3-5) Afrotheria is recognized in full to include also two families of endemic African insectivores, the tenrecs (Tenrecidae) and the golden moles (Chrysochloridae). This divides the traditional order Insectivora into at least two groups with different relationships. One group of African origin (Afrosoricida) and one of Laurasian origin (Eulipotyphla) (figure 1). This result was obtained by broadening the number of genetic markers as well as the number of taxa for each gene. As the name indicates, all members of Afrotheria have African origins. Molecular clock estimations indicate that the earliest divergences within the Afrotheria occurred some 70 to 80 million years ago, implying that an extensive radiation from a single common ancestor into ecologically divergent adaptive types took place during Africa's isolation in the Cretaceous (chapter 3). From the analyses of single gene fragments with limited length - up to about 1000 base pairs - it became clear that such data only gave consistent and robust support for a few superordinal clades, such as Afrotheria (chapters 2-5), and for the nesting of whales inside the Artiodactyla as sister to hippopotamus (see Gatesy and O'Leary 2001, for review). Thus, concatenating different genes into a single data matrix seemed to be a logical next step for gaining increased resolving power at the deeper level of the eutherian tree (chapter 6). This resulted, for the first time, in the resolution of eutherians into four major groups: Afrotheria, Xenarthra, Laurasiatheria and Euarchontoglires (figure 1). Additionally, a

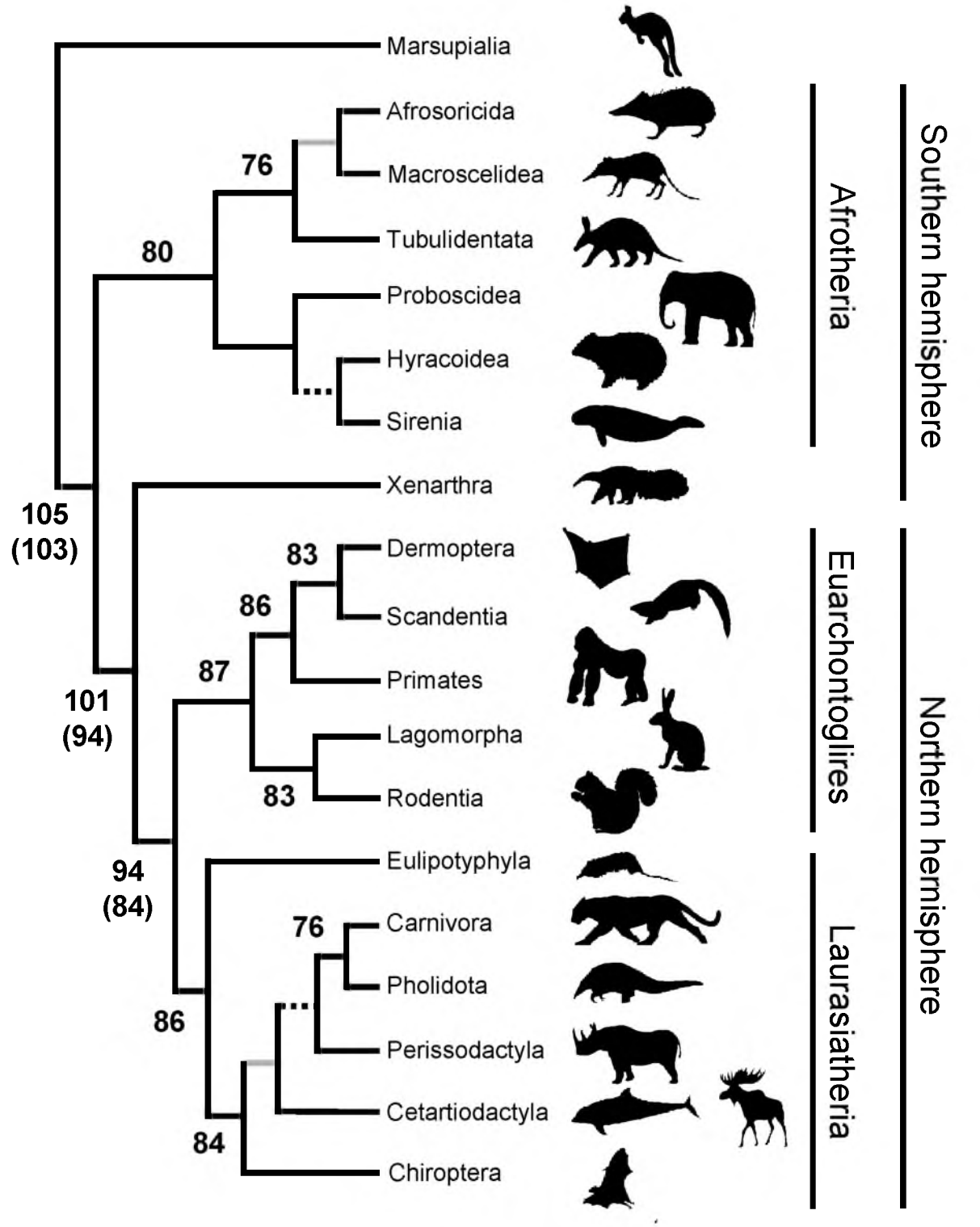


Figure 1. Phylogenetic tree of the orders of eutherian mammals as based on Murphy et al. (2001b). All nodes were robustly supported, except the dotted nodes, which were supported by Bayesian posterior probability and maximum likelihood non-parametric bootstrap values between 0.97-0.70 and under 50%, respectively, and the gray nodes, which were supported by Bayesian posterior probability and maximum likelihood non-parametric bootstrap values between 0.99-0.98 and between 50-75%, respectively. The numbers indicate molecular divergence times in million years ago, either from Murphy et al (2001b; in brackets), or from estimates on the Murphy et al. (2001b) data set with the method of Thorne et al. (1998) which allows a relaxed molecular clock and multiple calibration points (Eizirik et al. unpublished; without brackets). The three major superordinal clades are indicated, as well as the first occurrence in the fossil record of representatives of the various orders on the southern and northern hemisphere according to McKenna and Bell (1997).

unique nine-base-pair deletion in the BRCA1 gene provided further support for the naturalness of Afrotheria. A similar division of eutherian mammals into four groups was independently obtained by analyses of a completely different data set consisting of 18 orthologous gene segments (Murphy et al. 2001a; Eizirik et al. 2001).

This eutherian four-partitioning suggests that at the morphological level parallel adaptive radiations have occurred in the major clades (chapter 6; Scally et al. 2001), reminiscent of the well-known parallel adaptive radiations described between eutherian and marsupial mammals (Springer et al. 1997). The most notable morphological convergences within eutherians are those of various insectivores: Erinaceidae (hedgehogs) in Laurasiatheria versus *Echinops* (Madagascar "hedgehogs") in Afrotheria; Soricidae (shrews) in Laurasiatheria versus *Microgale* (shrewlike tenrecs) in Afrotheria; and Talpidae (moles) in Laurasiatheria versus Chrysochloridae (golden moles) in Afrotheria.

Although these studies on large combined data sets gave robust resolution for the four major groups, several relationships within these clades and the exact placement of the eutherian root remained still unresolved (see chapter 6). The position of the root is of utmost relevance for understanding the early radiation and biogeographic distribution of eutherian mammals. Recently, the data sets from chapter 6 and from Murphy et al. (2001a) have been combined and expanded, yielding a data matrix consisting of 22 gene segments of mostly nuclear origin, having a total length of 16.397 base pairs for 42 eutherian and two marsupial taxa (Murphy et al. 2001b). Phylogenetic relationships were assessed by using the traditional maximum-likelihood (ML) method as well as a Bayesian approach (MRBAYES; Huelsenbeck and Ronquist 2001) which is a new and powerful method for obtaining phylogenetic trees. Whereas the ML method searches a single ML tree, Bayesian methods sample and evaluate trees according to their posterior probability, given the data and the model of sequence evolution. A nearly completely resolved eutherian tree, the one shown in figure 1, was obtained by these analyses. Only two nodes remained ambiguously supported, and two others obtained medium to high support, whereas the rest was very robustly supported. At the base of the tree Afrotheria is the first to branch off, followed by Xenarthra, while Euarchontoglires and Laurasiatheria are found as sister clades. Comparison of this tree with the latest mitochondrially (MT) based tree (Arnason et al. 2002), shows many similarities but also a few differences. Most of the differences are related to the rooting of the eutherian tree, since unrooted analyses on the MT data result in the same topology as unrooted analyses on the nuclear data (O. Madsen, unpublished observation).

Combining the basal resolution of the eutherian tree with evidence from the fossil record, which indicates that Afrotheria and Xenarthra have first appearance in the southern hemisphere (Gondwana), while Euarchontoglires and Laurasiatheria first appeared in the northern hemisphere (Laurasia) (McKenna and Bell 1997) (see figure 1), a biogeographic scenario for the origin and early diversification of extant eutherian mammals can be suggested. Assuming that the first appearance of a taxon in the fossil record is likely to indicate its place of origin, the most parsimonious explanation for these observations is that extant eutherians had their last common ancestor in the southern hemisphere. This contradicts the traditional view that the last common ancestor of extant eutherians is of northern hemispheric origin (Wallace 1962; Rainger 1991). Molecular clock estimations date the divergence of Afrotheria from the other eutherians at around 105 million years (Myr) ago (see figure 1). It is notable that this dating is in good agreement with the continental breakup between South America and Africa, making vicariant separation a likely trigger for the basal divergence of eutherians, followed by a single trans-hemispheric dispersal from Gondwana to Laurasia (Murphy et al. 2001b).

The traditional paleontological view of eutherian divergence is that starting from a few eutherian lineages a rapid and bushlike radiation took place around 65 Myr ago, after the dinosaurs went extinct. However, estimations of the molecular divergence times of the different eutherian orders, as depicted in figure 1 (Murphy et al. 2001b; Eizirik et al. unpublished), go much deeper in time. These estimates are in line with other molecular estimates (e.g. Kumar and Hedges 1998; Penny et al. 1999) and indicate an expanded diversity of mammals long before the dinosaurs went extinct, with almost all extant orders already present at the Cretaceous/Tertiary (K/T) boundary. Hence it follows that a gap in the fossil record is likely to be present (Hedges and Kumar 1999). There are at least two possible explanations for such a gap. 1) Genetic isolation, and consequently the estimated molecular divergence times, will normally precede any observed synapomorphous morphological characters that can distinguish between two genetically isolated populations. Thus, some early fossils may actually be ancestors to present-day orders, but without being recognized as such, because of the absence of any descendant-specific derived characters (Easteal 1999; see also chapter 6). 2) The relatively poor fossil record from the southern hemisphere continents (especially from Africa) together with the possible Gondwanan origin of modern eutherians may suggest that these early fossils have not yet been found (Foote et al. 1999).

Establishing a well-supported eutherian tree, reflecting the actual course of divergent events, is extremely useful for understanding several different aspects of mammalian evolution. Firstly, it can be used as a scaffold to pinpoint the mode of morphological evolution in eutherians. Such an approach has already been used to study the evolution of laryngeal echolocation in bats (Springer et al. 2001). The molecular phylogenetic tree of bats was used as a scaffold to establish that laryngeal echolocation was present in ancestral bats and secondarily lost in megabats. Secondly, the current tree places rodents and primates as fairly close relatives, which at first sight supports the choice of mouse and rat as model animals for studying human diseases. But, it should be noted that these murine rodents display a rather distinct molecular substitution rate as compared to many other rodents and mammals (see tree in chapter 6; Eizirik et al. 2001; Waddell et al. 2001) indicating deviant cellular mechanisms in murines. Thus, instead of using rat and mouse as model animals for human diseases, a rodent such as squirrel or guinea pig had probably been a much better choice. Thirdly, in mammalian genome research the tree can help to choose the most appropriate new species for further genome sequencing projects. The mammals with (nearly) completely sequenced genomes (human, rat and mouse) all belong to the same phylogenetic clade. To better comprehend mammalian genome evolution, the genomes of several different and preferably phylogenetically most divergent species from the other three eutherian clades plus outgroup mammals should be sequenced. Besides a better understanding of mammalian genome evolution, the analysis of these sequences will of course provide the final validation of the eutherian tree.

A last and most important phylogenetic question is of course whether the present well-resolved eutherian tree is tantamount to 'we are finished' with concern to establishing eutherian relationships at the ordinal level as based on molecular sequence data. The answer to this question is yes and no. **Yes**, because determining ever longer sequences and more extended species sampling is not likely to result in another tree topology or to resolve the ambiguous parts of the tree. At least not with the currently used models of sequence evolution. Although some models of sequence evolution are fairly complex, they still do not describe most data in full (Goldman 1993). Thus, by including models and parameters such as codon models (Muse and Gaut 1994; Yang et al. 2000), within-site rate

variation parameters (heterotachy; Lopez et al. 2002), models based on structural information (Topham et al. 1993; Lió and Goldman, 2002), and covarion model parameters (e.g. Huelsenbeck, 2002), better descriptions of the data are likely to be achieved, and may result in an even better resolved eutherian tree. **No**, because independent confirmation of the proposed eutherian relationships from other genetic sources such as rare genomic events (Rokas and Holland 2000) is very important. Not only will such independent data give better confidence in the tree, they also may help to convince any sceptical mammalian systematists that the current molecular tree best approaches the true tree. Rare genomic changes are for example insertions or deletions in protein-coding genes (indels), alternative splicing, and short interspersed elements (SINEs). In eutherian systematics such rare genomic changes have already been used as phylogenetic markers: indels have been found supporting Afrotheria (chapter 6), Xenarthra (van Dijk et al. 1999) and Euarchontoglires (Poux et al. 2002); and SINEs have been used in resolving cetartiodactyl and primate intraordinal relationships (e.g. Nikaido et al. 1999; Schmitz et al. 2001).

In conclusion, a well-resolved eutherian tree has been obtained, but there is still work to do in terms of improving the models and solving the last remaining ambiguities.

Molecular Evolution

In addition to the phylogenetic results, what can we learn about molecular evolution from the many kilobases of gene sequences that have been determined in the course of this phylogenetic study? In molecular databases such as EMBL and Genbank the great majority of mammalian sequences is from human, mouse and rat. This bias has become even more pronounced in this postgenomic era with almost complete genome sequences of human, mouse and rat. Therefore, if a human gene is described with unknown or new structures and functions, it often is compared with its mouse or rat counterpart to pinpoint important regions in the protein for further investigations. Since primates and rodents are fairly closely related mammals (figure 1), and parallel substitutions can not be detected when aligning only two sequences, it is unlikely that such a comparison is able to unveil all of the important regions in the protein. This limitation advocates a broader sampling of mammalian sequences, which will provide a much more comprehensive comparison of the variability and conservation of a gene or protein.

In this study two nuclear encoded mammalian genes from a broad range of different mammals have been analysed: exon 1 of aquaporin 2 (AQP2) which is a member of the AQP family (chapter 2), and the alpha 2B adrenergic receptor (A2AB) which is a class A receptor of the G protein-coupled receptor (GPCR) family (chapter 7). Both genes code for transmembrane (TM) proteins but with entirely different functions, AQP2 facilitating active transportation of water across the cell membrane, and A2AB involved in signal transmission from the extra- to the intracellular environment. As expected, residues and domains predicted to be structurally and functionally important in the two proteins, such as the aqueous pore domain (NPA domain) in AQP2, and residues involved in ligand- and G protein-coupling in A2AB, are highly conserved. In general, TM domains are more conserved than the loops that connect these domains, implying lower functional and structural constraints on loops than on the TM domains. However, different loops may show great variation in conservation, from well-conserved as for intracellular loop 1 in A2AB, to highly variable as for intracellular loop 3 in A2AB. In the latter loop 3 only a few residues and a polyglutamyl domain are preserved. Although

this acid domain displays great length variation, it is present in all described mammalian A2ABs, confirming its proposed role in agonist-dependent phosphorylation of the third intracellular loop (Jewell-Motz et al. 1995; Small et al. 2001). The length variation in the acid domain might indicate that the level of agonist-dependent phosphorylation and consequent desensitization and surface receptor activity may be regulated by an evolutionary fast process required for adaptation to changing cellular and environmental demands in different species. It further illustrates that amino acid repeats could play an important role in protein function, as has been suggested for other proteins (Karlin and Burge 1996).

Recently, the crystal structures of rhodopsin, an A2AB-related class A GPCR, and of AQP1 which is related to AQP2, have been resolved (Murata et al. 2000; Palczewski et al. 2000). The conservation and variation in AQP2 and A2AB as observed from the multiple alignments of these proteins in our studies, are more or less consistent with the high-resolution structures of AQP1 and rhodopsin. This agrees with their structural relatedness, but there are differences between our deductions for AQP2 and A2AB, and the observations from the AQP1 and rhodopsin crystal structures. For example, in rhodopsin two antiparallel β-strands are present in the second extracellular loop, of which the latter forms part of the ligand-binding pocket (Palczewski et al. 2000). The corresponding regions in A2AB are quite variable, suggesting that A2AB and rhodopsin have different structures in this part of the molecule (chapter 7). Additionally, recent studies have indicated that the secondary structure of the second intracellular loop, which plays an important role in regulating G protein-coupling, differs between the alpha 2A adrenergic receptor (A2AA) and rhodopsin (Chung et al. 2002a), and that in the class A family of GPCRs at least two different G protein-coupling mechanisms exist (Lu et al. 1997; Burstein et al. 1998; Chung et al. 2002b). Thus, the high-resolution structure of rhodopsin provides an excellent template for other class A GPCRs in establishing the structures of the seven TM bundles (Bourne and Meng, 2000), but the structures of the different intracellular and extracellular loops and their exact involvement in determining ligand-binding and G protein-coupling still need to be established.

In conclusion, determining and comparing in a broad range of mammals the sequences of a particular orthologous protein with unknown function or structure, seems to be a good first step for gaining more knowledge about the most important parts of that protein. This information then can subsequently be used for further investigations, such as making point or deletion mutants of the protein and performing protein structure predictions.

References:

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, and Janke A. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc. Natl. Acad. Sci. USA* **99**: 8151-8156

Bourne HR, and Meng EC. 2000. Structure. Rhodopsin sees the light. *Science* **289**: 733-734

Burstein ES, Spalding TA, and Brann MR. 1998. The second intra-cellular loop of the m5 muscarinic receptor is the switch which enables G-protein coupling. *J. Biol. Chem.* **273**: 24322-24327

Cao Y, Adachi J, Janke A, Paabo S, and Hasegawa M. 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial pro-teins: instability of a tree based on a single gene. *J. Mol. Evol.* **39**: 519-527

Chung DA, Zuiderweg ER, Fowler CB, Soyer OS, Mosberg HI, and Neubig RR. 2002a. NMR structure of the second intracellular loop of the alpha 2A adrenergic receptor: evidence for a novel cytoplasmic helix. *Biochemistry* **41**: 3596-3604

Chung DA, Wade SM, Fowler CB, Woods DD, Abada PB, Mosberg HI, and Neubig RR. 2002b. Mutagenesis and peptide analysis of the DRY motif in the alpha2A adrenergic receptor: evidence for alternate mechanisms in G protein-coupled receptors. *Biochem. Biophys. Res. Commun.* **293**: 1233-1241

de Jong WW. 1998. Molecules remodel the mammalian tree. *Trends Ecol. Evol.* **13**: 270-275
Easteal S. 1999. Molecular evidence for the early divergence of placental mammals. *Bioessays* **21**: 1052-1058

Eizirik E, Murphy WJ, and O'Brien SJ. 2001. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* **92**: 212-219

Foote M, Hunter JP, Janis CM, and Sepkoski JJ Jr. 1999. Evolutionary and preservational constraints on origins of biologic groups: Diver-gence times of eutherian mammals. *Science* **283**: 1310-1314

Gatesy J, and O'Leary MA. 2001. Deciphering whale origins with molecules and fossils. *Trends Ecol. Evol.* **16**: 562-570

Goldman N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36**: 182-198

Hedges SB, and Kumar S. 1999. Divergence times of eutherian mammals. *Science* **285**: Technical comments 2031a

Honeycutt RL, Nedbal MA, Adkins RM, and Janecek LL. 1995. Mammalian mitochondrial DNA evolution: a comparison of the cytochrome b and cytochrome c oxidase II genes. *J. Mol. Evol.* **40**: 260-272

Huelsenbeck JP, and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755

Huelsenbeck JP. 2002. Testing a covariotide model of DNA substitution. *Mol. Biol. Evol.* **19**: 698-707

Jewell-Motz EA, and Liggett SB. 1995. An acidic motif within the third intracellular loop of the alpha2C2 adrenergic receptor is required for agonist-promoted phosphorylation and desensitization. *Biochemistry* **34**: 11946-11953

Karlin S, and Burge C. 1996. Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development. *Proc. Natl. Acad. Sci. USA* **93**: 1560-1565

Kumar S, and Hedges SB. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**: 917-920

Li WH, Gouy M, Sharp PM, O'hUigin C, and Yang YW. 1990. Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. *Proc. Natl. Acad. Sci. USA* **87**: 6703-6707

Liò P, and Goldman N. 2002. Modeling mitochondrial protein evolution using structural information. *J. Mol. Evol.* **54**: 519-529

Lopez P, Casane D, and Philippe H. 2002. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* **19**: 1-7

Lu ZL, Curtis CA, Jones PG, Pavia J, and Hulme EC. 1997. The role of the aspartate-arginine-tyrosine triad in the m1 muscarinic receptor: mutations of aspartate 122 and tyrosine 124 decrease receptor expression but do not abolish signaling. *Mol. Pharmacol.* **51**: 234-241

McKenna, MC. and Bell, SK. 1997. Classification of Mammals Above the Species Level, Columbia Univ. Press, New York

Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann JB, Engel A, and Fujiyoshi Y. 2000. Structural determinants of water permeation through aquaporin-1. *Nature* **407**: 599-605

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, and O'Brien SJ. 2001a. Molecular phylogenetics and the origins of placental mammals. *Nature* **409**: 614–618

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, and Springer MS. 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**: 2348-2351

Muse SV, and Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* **11**: 715-724

Nikaido M, Kawai K, Cao Y, Harada M, Tomita S, Okada N, and Hasegawa M. 2001. Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* **53**: 508-516

Penny D, Hasegawa M, Waddell PJ, and Hendy MD. 1999. Mammalian evolution: Timing and implications from using the LogDeterminant transform for proteins of differing amino acid

composition. *Syst. Biol.* **48**: 76-93

Palczewski K, Kumasaka T, Hori T, et al., (12 co-authors). 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**: 739-745

Porter CA, Goodman M, and Stanhope MJ. 1996. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gene. *Mol. Phylogenet. Evol.* **5**: 89-101

Poux C, van Rheede T, Madsen O, and de Jong WW. 2002. Sequence Gaps Join Mice and Men: Phylogenetic Evidence from Deletions in Two Proteins. *Mol. Biol. Evol.* (in press)

Rokas A, and Holland PWH. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* **15**: 454-459

Rainger R. 1991. *Agenda for Antiquity: Henry Fairfield Osborn and Vertebrate Paleontology at the American Museum of Natural History, 1890-1935*, University of Alabama Press, Tuscaloosa, Alabama

Scally M, Madsen O, Douady CJ, de Jong WW, Stanhope, MJ and Springer MS. 2001. Molecular Evidence for the Major Clades of Placental Mammals. J. Mammal. Evol. 8: 239-277.

Schmitz J, Ohme M, and Zischler H. 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of Tarsius bancanus to other primates. *Genetics* **157**: 777-784

Small KM, Brown KM, Forbes SL, and Liggett SB. 2001. Polymorphic deletion of three intracellular acidic residues of the alpha 2B-adrenergic receptor decreases G protein-coupled receptor kinase-mediated phosphorylation and desensitization. *J. Biol. Chem.* **267**: 4917-4922

Springer MS, Kirsch JAW, and Case JA. 1997. The chronicle of marsupial evolution. In: *Molecular Evolution and Adaptive Radiation*, T. J. Givnish and K. J. Sytsma, eds., pp. 129-161, Cambridge University Press, New York.

Springer MS, Teeling EC, Madsen O, Stanhope MJ, and de Jong WW. 2001. Integrated fossil and molecular data reconstruct bat echolocation. *Proc. Natl. Acad. Sci. USA* **98**: 6241-6246

Stanhope MJ, Smith MR, Waddell VG, Porter CA, Shivji MS, and Goodman M. 1996. Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. *J. Mol. Evol.* **43**: 83-92

Thorne JL, Kishino H, and Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**: 1647-1657

Topham CM, McLeod A, Eisenmenger F, Overington JP, Johnson MS, and Blundell TL. 1993. Fragment ranking in modelling of protein structure. Conformationally constrained environmental amino acid substitution tables. *J. Mol. Biol.* **229**: 194-220

van Dijk MA, Paradis E, Catzeflis F, and de Jong WW. 1999. The virtues of gaps: xenarthran (Edentate) monophyly supported by a unique deletion in alpha A-crystallin. *Syst Biol* **48**: 94-106

Waddell PJ, Kishino H, and Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Inform.***12**: 141-54

Wallace AR. 1962. *The Geographical Distribution of Animals*, 2 volumes, Hafner, New York.

Yang Z, Nielsen R, Goldman N, and Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431-44

Het in dit proefschrift beschreven onderzoek diende een tweeledig doel. Allereerst het verkrijgen van een dieper inzicht in de fylogenetische verwantschappen tussen de hedendaagse zoogdieren door onderzoek van uitgebreide bestanden van moleculaire gegevens. Daarnaast het beschrijven van de moleculaire evolutie van de genen die voor deze fylogenetische studies gebruikt zijn, in samenhang met hun structuren en functies.

In dit onderzoek zijn twee genen uit het kerngenoom vergeleken bij een breed scala van zoogdieren: exon 1 van het gen dat codeert voor aquaporine 2 (AQP2), een lid van de AQP familie (hoofdstuk 2) en het gen dat codeert voor de alfa 2B adrenerge receptor (A2AB), een klasse A receptor uit de familie van G-eiwit gekoppelde receptoren (GPCRs; hoofdstuk 7). Beide genen coderen voor transmembraan (TM) eiwitten, doch met volstrekt verschillende functies. AQP2 maakt het actieve transport van water over de celmembraan mogelijk, terwijl A2AB betrokken is bij de overdracht van signalen vanuit de extracellulaire naar de intracellulaire omgeving. Zoals verwacht bleken vooral die aminozuren en eiwitdomeinen evolutionair sterk geconserveerd waarvan voorspeld wordt dat ze structureel en functioneel van belang zijn voor de betreffende eiwitten. Voor AQP2 geldt dat met name voor het domein rond de waterporie, gekenmerkt door de drie aminozuren asparagine, proline en alanine (het NPA domein). Voor A2AB betreft dat vooral de aminozuren die betrokken zijn bij de binding van liganden en G-eiwitten. Over het algemeen blijken TM domeinen sterker geconserveerd te zijn dan de lussen die deze domeinen verbinden, wat er op wijst dat de lussen functioneel en structureel aan minder beperkingen onderhevig zijn dan de domeinen. De lussen vertonen onderling echter grote verschillen in conservering van hun aminozuurvolgorden, in A2AB uiteenlopend van een sterk geconserveerde intracellulaire lus 1 tot een zeer variabele intracellulaire lus 3. Geconcludeerd kan worden dat het bepalen en vergelijken van de aminozuurvolgorden van een overeenkomstig eiwit bij een goed gekozen selectie van zoogdieren een goede indruk geeft van de meest belangrijke delen van zo'n eiwit, ook wanneer de functie en structuur nog onbekend zijn. Deze informatie kan vervolgens in verdere experimenten benut en getest worden.

De hoofdstukken 2 tot en met 6 beschrijven opeenvolgende stappen in het ontrafelen van de verwantschappen tussen de verschillende orden van placentale zoogdieren, de Eutheria, op basis van de vergelijking van DNA volgorden van hun genen. Dit begint in hoofdstuk 2 met de bevestiging dat aardvarkens en olifantspitsmuizen verwant zijn met zeekoeien, olifanten en klipdassen, en samen een groep van Afrikaanse oorsprong vormen, die thans als Afrotheria bekend staan. In de hoofdstukken 3-5 wordt dan beschreven dat ook twee families van endemische Afrikaanse insecteneters, de goudmollen en de tenreks, tot de Afrotheria behoren. Aangezien goudmollen en tenreks tot nu toe samen met egels, mollen en spitsmuizen in de orde Insectivora geplaatst werden, moet geconcludeerd worden dat deze orde ten minste in twee groepen verdeeld moet worden, een van Afrikaanse oorsprong en een met oorsprong op de noordelijke continenten, Laurazië. In hoofdstuk 3 wordt op grond van de volgordeverschillen tussen de zoogdiergenen, en het gegeven dat DNA ruwweg met constante snelheid evolueert (de "moleculaire klok"), gesuggereerd dat de verschillende orden binnen de Afrotheria tussen midden Krijt en vroeg Kaenozoicum uit een gemeenschappelijke voorouder gedivergeerd zijn. Dit was de periode dat Afrika zich reeds van Zuid-Amerika had afgescheiden maar nog geen contact had met Eurazië. Ten slotte beschrijft hoofdstuk 6 hoe de placentale zoogdieren in vier hoofdgroepen kunnen worden verdeeld, waarvan de Afrotheria er één vormen. Deze verdeling kon worden vastgesteld door de volgorden van een aantal afzonderlijke genen te combineren in één grote matrix van gegevens en daarna te analyseren. Op morfologisch vlak hebben zich binnen de vier

hoofdgroepen parallelle en convergente adaptieve veranderingen voorgedaan. Het meest opvallend is de morfologische convergentie bij de verschillende insecteneters binnen twee van de hoofdgroepen, de Afrotheria en de Laurasiatheria. Zo lijken de tenreks van de geslachten *Echinops* ("Madagascar hedgehogs") en *Microgale* ("shrewlike tenrecs") en de Afrikaanse goudmollen, alle Afrotheria, zoals de namen al aangeven sprekend op onze Euraziatische egels, spitsmuizen en mollen, die tot de Laurasiatheria behoren.

Het beschreven onderzoek heeft er toe bijgedragen dat de stamboom van de zoogdieren thans voor een belangrijk deel is opgehelderd. Deze boom, gebaseerd op moleculaire gegevens, vertoont een aantal verrassende verwantschappen die niet uit eerdere paleontologische en morfologische stambomen naar voren gekomen waren.

Ph.D.-projektet, som er beskrevet i denne afhandling havde to formål. For det første at opnå et bedre kendskab til klassifikationen af nulevende pattedyr ud fra omfattende DNA sekvens data. For det andet at beskrive den molekylære evolution (udvikling) i forhold til deres struktur og funktion af de kerne-DNA gener, som blev anvendt i disse fylogenetiske studier.

Den molekylære evolution af to kerne-kodende gener: exon 1 af vandkanalprotein 2 (AQP2), som tilhører familien af vandkanalproteiner og alfa 2B adrenergene receptor (A2AB), som tilhører familien af G-protein koblede receptorer, er beskrevet i henholdsvis kapitel 2 og 7. Begge gener koder for transmembrane (TM) proteiner, som dog har vidt forskellige funktioner. AQP2 spiller en vigtig rolle i den aktive vandtransport over cellemembranen i nyrerne, mens A2AB er involveret i overførslen af signaler fra det extra- (udenfor cellen) til det intracellulære (indeni cellen) miljø. Den molekylære evolution af disse to gener blev undersøgt ved at sammenligne protein sekvenser, fra mange forskellige grupper af pattedyr, hvorved det var muligt at identificere de vigtigste domæner og aminosyrer i de to proteiner, eftersom disse domæner og aminosyrer er de bedst konserverede dele af et protein. Som forventet er aminosyrerne i og omkring protein domænet, der er ansvarlig for den aktive vandtransport over cellemembranen, vel konserveret i alle AQP2 molekylærene. I A2AB er aminosyrerne og protein domænerne involveret i henholdsvis ligand og G-protein binding også godt konserveret. I almindelighed er TM domænerne bedre konserveret end protein løkkerne, som forbinder disse TM domæner, hvilket tyder på, at der er en større evolutionær begrænsning på TM domænerne end på protein løkkerne. Der er imidlertid stor forskel i konserveringsgraden af de forskellige løkker, vekselende fra meget konserveret (løkke 1, A2AB) til meget variabel (løkke 3, A2AB).
Sammenfattende kan det konkluderes, at en god metode til at identificere de vigtigste områder med hensyn til struktur og funktion i et protein er sammenligning af protein sekvenser fra mange forskellige dyr/organismer.

I kapitlerne 2 til og med 6 beskrives den gradvise forbedring i etableringen af det fylogenetiske slægtskab imellem de forskellige ordener af placentale (eutherian) pattedyr. Dette forbedrede stamtræ blev opnået ved at sammenligne forskellige gen sekvenser fra de forskellige placentale pattedyr. I kapitel 2 bekræftes det, at jordsvin og springspidsmus er nært beslægtede med søkøer, elefanter og klippegrævlinger. Fællesnævneren for disse morfologiske meget forskellige dyr er, at de alle har en afrikansk oprindelse og derfor blev de kaldt den "Afrikanske gruppe"; en gruppe der senere blev døbt Afrotheria. I de følgende kapitler (3 til 5) er det beskrevet, at to familier af endemiske afrikanske insektædere, børstesvin og guldmuldvarpe, også tilhører Afrotheria gruppen. Eftersom børstesvin og guldmuldvarpe normalt, sammen med spidsmus, muldvarpe og pindsvin er inddelt i ordenen Insectivora, kan det konkluderes, at denne orden i det mindste skal splittes op i to grupper, en med afrikansk oprindelse og en med oprindelse på den nordlige halvkugle, Laurasia. Ved at gå ud fra, at ankomsttiderne af DNA-mutationer er jævnt fordelt over tid, det såkaldt "molekylære ur", var det muligt at beregne, at de forskellige udviklingslinier fra den sidste fælles stamfader i Afrotheria fandt sted i tidsrummet mellem kridt indtil tidlig kænozoisktid, hvor Afrika var en isoleret ø. Til sidst i kapitel 6 beskrives det, for første gang, at placentale pattedyr kan deles op i 4 hovedgrupper, hvoraf Afrotheria er en af disse hovedgrupper. (Dette resultat blev opnået ved at kombinere en række forskellige gen-fragmenter i et fælles molekylær datasæt.) I de forskellige hovedgrupper har en række forskellige konvergente og parallelle morfologiske tilpasninger fundet sted. De mest iøjnefaldende konvergente tilpasninger findes iblandt insektæderne i de to hovedgrupper

Afrotheria og Laurasiatheria, så som: Madagaskar børstesvin ("pindsvin") (*Echinopus*) versus pindsvin (Erinaceidae) og guldmuldvarpe (Chrysochloridae) versus muldvarpe (Talpidae).
Således er der ved hjælp af molekylære data opnået et veldefineret stamtræ for placentale pattedyr, som angiver nye spændende fylogenetiske slægtskaber, der ikke tidligere har været foreslået baseret på morfologiske eller palæontologiske data.

Madsen O, Sandal L, Sandal NN, Marcker KA. (1993) A soybean coproporphyrinogen oxidase gene is highly expressed in root nodules. *Plant. Mol. Biol.* 23:35-43

Madsen O, Deen PM, Pesole G, Saccone C, de Jong WW. (1997) Molecular evolution of mammalian aquaporin-2: further evidence that elephant shrew and aardvark join the paenungulate clade. *Mol. Biol. Evol.* 14:363-371

Springer MS, Cleven GC, Madsen O, de Jong WW, Waddell VG, Amrine HM, Stanhope MJ. (1997) Endemic African mammals shake the phylogenetic tree. *Nature* 388:61-64

Stanhope MJ, Madsen O, Waddell VG, Cleven GC, de Jong WW, Springer MS. (1998) Highly congruent molecular support for a diverse superordinal clade of endemic African mammals. *Mol. Phylogenet. Evol.* 9:501-508

Stanhope MJ, Waddell VG, Madsen O, de Jong W, Hedges SB, Cleven GC, Kao D, Springer MS. (1998) Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl. Acad. Sci. USA* 95:9967-9972

van Dijk MA, Madsen O, Catzeflis F, Stanhope MJ, de Jong WW, Pagel M. (2001) Protein sequence signatures support the African clade of mammals. *Proc. Natl. Acad. Sci. USA* 98:188-193

Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, de Jong WW, Stanhope MJ. (2001) Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* 18:132-143

Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM, Stanhope MJ, de Jong WW, Springer MS. (2001) Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409:610-614

Springer MS, Teeling EC, Madsen O, Stanhope MJ, de Jong WW. (2001) Integrated fossil and molecular data reconstruct bat echolocation. *Proc. Natl. Acad. Sci. USA* 98:6241-6246

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS. (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348-2351

Scally, M., O. Madsen, C.J. Douady, W.W. de Jong, M.J. Stanhope, and M.S. Springer. (2001) Molecular Evidence for the Major Clades of Placental Mammals. *J. Mammal. Evol.* 8:239-277.

Teeling EC, Madsen O, Van den Bussche RA, de Jong WW, Stanhope MJ, Springer MS. (2002) Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. *Proc. Natl. Acad. Sci. USA* 99:1431-1436

Huchon D, Madsen O, Sibbald MJ, Ament K, Stanhope MJ, Catzeflis F, de Jong WW, Douzery EJ. (2002) Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. *Mol. Biol. Evol.* 19:1053-1065

Delsuc F, Scally M, Madsen O, Stanhope MJ, de Jong WW, Catzeflis FM, Springer MS, Douzery EJ. (2002) Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Mol. Biol. Evol.* 19:1656-1671

Douady C, Chatelier P, Madsen O, de Jong W, Catzeflis F, Springer M, Stanhope M. (2002) Molecular phylogenetic evidence confirming the Eulipotyphla concept and in support of hedgehogs as the sister group to shrews. *Mol. Phylogenet. Evol.* 25:200-209

Madsen O, Willemsen D, Ursing BM, Arnason U, de Jong WW. (2002) Molecular Evolution of the Mammalian Alpha 2B Adrenergic Receptor. *Mol. Biol. Evol. In press*

Poux C, van Rheede T, Madsen O, de Jong WW. (2002) Sequence Gaps Join Mice and Men: Phylogenetic Evidence from Deletions in Two Proteins *Mol. Biol. Evol. In press*

van Rheede T, Smolenaars M, Madsen O, de Jong WW. (2002) Molecular evolution of the mammalian prion proteins. *Mol. Biol. Evol. In press*

Ole Madsen was born the twelve of September 1965 in Ringkøbing (Denmark)

Education:

1981-1984	Ringkøbing Amtsgymnasium, Gymnasium-degree, Nature science-line
1986-1993	Aarhus University, Biology (M.Sc.), specialization molecular biology and biochemistry
1994-1999	University of Nijmegen, Ph.D. student, Molecular phylogeny of mammals

Scientifically work experience:

1998	Queen's University of Belfast, Research Associate
2000-2003	University of Nijmegen, Post Doc, Mammalian phylogeny

Nothing in this world pays for itself and so is this study. A lot of people have supported me during my Ph.D. study. A great thanks to all these people. A few of them I would like to mention by name.

First of all I would like to thank my supervisor Wilfried de Jong. Wilfried, when I started my Ph.D. study I knew little about evolution and even less about mammalian systematic, but your enthusiasm for evolution and systematic has an infectious influence and is always a constant well of inspiration for me. You also helped me a lot with finishing this book and improving several chapters. Also, I would like to thank you for your emancipated attitude when it comes to the needs of fathers with respect to flexible working hours.

My collaborators also get a great thanks. Especially the co-operation with Mark Springer and Michael Stanhope was very fruitful and educational. Without the co-operation with Mark, Michael and their co-workers the scientifically impact of this book would probably have been a magnitude lower. Michael, too, get thanks for employing me as research associate for six and a half month. Also the European network meetings were worth-while experiences and inspiring. In particular the night meetings with the other "young" scientists learned me a lot about other cultures. May the "blue frog" stay with us.

I am also very thankful to my colleagues in the lab. Marjon, as you wrote in your thesis, for long we formed the phylogenetic branch in the lab. It was always a pleasure working together with you, as we had a lot of good discussions and you were always very helpful in the lab. You have a wonderful spirit. Do not forget to keep on laughing as your laugh is very optimistic. Also thanks to the two new "phylogeneticist" Teun and Celine for many discussions. No computers no phylogenetic studies: Paul (Werten), you really taught me a lot about computers and deserve a lot of gratitude; Guido (KP), thanks for always answering all my stupid PC questions; Jack (Leunissen), thanks for helping me through the CMBI jungle. I would also like to thank my students Diederik, Natascha and Kai. Although the work of the latter two is not presented in this book they all helped me a lot and contributed to make me at better scientist. Els, je bent de enige die een dankwoord in het Nederlands krijgt. Bedankt voor al je hulp bij allerlei dingen. Het is altijd leuk om met jou te kletsen en als er meer mensen zoals jij gewoon zouden zeggen waar het op staat, zou deze wereld er veel beter uitzien. Thanks also to my various "Utje" colleagues, Pauline, Lisa, Bas and Marcel for their omnipresent interest in my work and family. Also thanks to Tinie for good friendship at and outside work.

Også mange tak til Bettina, Grethe og Trine for at I ville korrigere mit resumé.

My eldest daughters, Milou and Femke, also deserve a really great thanks for their beautiful drawings which can be admired elsewhere in this book and for keep asking WHEN my book was going to be finish. Now you probably are allowed to play Harry Potter more often on my computer. My youngest daughter Stine deserves a thanks for being such a sweet little baby. Last, but of course not least, I would like to thank my beloved wife Martine. Without her this thesis would never have existed.

Portrait of the author